



# Linguistic expectation management in online discourse processing: An investigation of Dutch *inderdaad* 'indeed' and *eigenlijk* 'actually'

Geertje van Bergen<sup>a,\*</sup>, Hans Rutger Bosker<sup>b</sup>

<sup>a</sup> Max Planck Institute for Psycholinguistics, Neurobiology of Language Department, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

<sup>b</sup> Max Planck Institute for Psycholinguistics, Psychology of Language Department, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

## ARTICLE INFO

### Keywords:

Discourse particles  
Discourse processing  
Dialogue  
Polyfunctionality  
Visual world eye-tracking

## ABSTRACT

Interpersonal discourse particles (DPs), such as Dutch *inderdaad* (≈‘indeed’) and *eigenlijk* (≈‘actually’) are highly frequent in everyday conversational interaction. Despite extensive theoretical descriptions of their polyfunctionality, little is known about how they are used by language comprehenders. In two visual world eye-tracking experiments involving an online dialogue completion task, we asked to what extent *inderdaad*, confirming an inferred expectation, and *eigenlijk*, contrasting with an inferred expectation, influence real-time understanding of dialogues. Answers in the dialogues contained a DP or a control adverb, and a critical discourse referent was replaced by a beep; participants chose the most likely dialogue completion by clicking on one of four referents in a display. Results show that listeners make rapid and fine-grained situation-specific inferences about the use of DPs, modulating their expectations about how the dialogue will unfold. Findings further specify and constrain theories about the conversation-managing function and polyfunctionality of DPs.

## Introduction

Some languages are equipped with specific linguistic tools to respond to contextual expectations, as illustrated in the following constructed Dutch dialogue:

- (1) A *Het was vast prachtig weer in Griekenland?*  
The weather must have been great in Greece?  
B *We hebben inderdaad alleen maar zon gehad.*  
We have [indeed] had nothing but sun.  
B' *We hebben eigenlijk alleen maar regen gehad.*  
We have [actually] had nothing but rain.

In these utterances, *inderdaad* (≈‘indeed’) and *eigenlijk* (≈‘actually, in fact’) respond to the expectation raised by A’s suggestive question: they mark either a match (*inderdaad*) or a mismatch (*eigenlijk*) between what B says and what A expects B to say. The current study investigates effects of such linguistic expectation-managing devices on online discourse comprehension.

### Expectation-managing discourse particles

*Inderdaad* and *eigenlijk* are discourse particles (DPs), a subset of the broad category of discourse markers. Discourse markers are linguistic elements that encode a relation between their host utterance and the

surrounding discourse situation (e.g., Fischer, 2006; Fraser, 1999; Schiffrin, 1987), and typically occur in highly dynamic, interactive contexts. They are syntactically flexible, in the sense that they can fall both within and outside the grammatical structure of the sentence (e.g., *we hebben alleen maar zon gehad, inderdaad ‘we have had nothing but sun, [indeed]’*). Moreover, they are syntactically optional, such that they can be left out of a sentence without violating its grammaticality (for instance, the answers in (1) would be syntactically well-formed without the DPs). Because of their extra- or a-grammatical status, discourse markers have long been regarded as meaningless verbal fillers not warranting linguistic investigation. However, a large body of research over the last three decades has revealed that they are “communicatively obligatory” (Diewald, 2010), and that their use requires sophisticated communicative knowledge (for reviews, see e.g., Degand, Cornillie, & Pietrandrea, 2013; Fischer, 2014; Fraser, 1999; Maschler & Schiffrin, 2015; Schiffrin, 1987).

From a semantics/pragmatics perspective, *inderdaad* and *eigenlijk* can be analyzed as conventional implicature triggers (e.g., Grice, 1975; Levinson, 1979; Potts, 2007) or “invited inferences” (Traugott & Dasher, 2002). Like presuppositions and entailments, conventional implicatures are inferences that are part of a lexical or constructional meaning. However, in contrast to presuppositions and entailments, they are independent from the at-issue content of the utterance (“what is said”) (for discussion, see Potts, 2015). Conventional implicatures are

\* Corresponding author.

E-mail addresses: [Geertje.vanBergen@mpi.nl](mailto:Geertje.vanBergen@mpi.nl) (G. van Bergen), [HansRutger.Bosker@mpi.nl](mailto:HansRutger.Bosker@mpi.nl) (H.R. Bosker).

highly context-dependent: the particular meaning they express is highly variable, and strongly influenced by the speaker's communicative goal as well as the nature of the surrounding discourse content (Potts, 2015, p. 30). Given this context-dependent nature of conventional implicatures, *inderdaad* and *eigenlijk* are notoriously polyfunctional: they can express multiple relations with linguistic, structural, cognitive and/or social characteristics of the discourse (e.g., Fischer, 2014; Maschler & Schiffrin, 2015). For instance, in example (1) above, *inderdaad* and *eigenlijk* mark a relation between the utterance in which they occur and some expectation that can be inferred from the question. Such an expectation can be based on various discourse aspects, e.g., the visual context (e.g., if one looks tanned, one has likely been exposed to a lot of sun), world knowledge (if one goes to Greece, one may expect good weather), but also the linguistic or the social context.

At the same time, DPs can apply to various levels of information expressed in their host utterance. For instance, if a child asks his mother for a piece of *kommer* “cumber” and his mother replies *het is eigenlijk “komkommer”* (it is [actually] “cucumber”), she uses *eigenlijk* to mark a mismatch between her own knowledge and her child's expectation about this word's pronunciation that can be inferred from the child's question. Another example: if two people are greatly enjoying themselves at a party, and one of them asks *hoe laat is het eigenlijk?* (what time is it [actually]?), *eigenlijk* is used to mark a contrast between the questioner's speech act (asking for the time) and the inference that you don't keep track of time if you are having fun (van Bergen, van Gijn, Hogeweg, & Lestrade, 2011). The interlocutor could respond to this question by saying *we moeten inderdaad maar eens naar huis gaan* (‘We should [indeed] head home’), *inderdaad* marking a match between the responder's intention to go home and the communicative intention of the questioner that can be implied from the speech act. In the theoretical literature, there is general agreement that all different interpretations of an individual discourse marker share a common core meaning (also referred to as *generic*, *abstract*, *underspecified* or *underlying* meaning; see e.g., Fischer, 2006; Fox Tree & Schrock, 2002). Similarly, we assume a high degree of relatedness between all different interpretations of *inderdaad* and *eigenlijk*, respectively: the former commonly marking affirmative/confirmatory discourse relations, the latter generally marking adversative/contrastive discourse relations (see also van Bergen et al., 2011; Mortier & Degand, 2009).

Together with this array of discourse-structuring functions, *inderdaad* and *eigenlijk* crucially express *interpersonal* or *intersubjective* relations: they manage the cognitive and/or social coordination between the speaker and the addressee or addressee's face (e.g., Traugott, 2010; Verhagen, 2005). In this sense, *inderdaad* and *eigenlijk* encode a relation between the (assumed) discourse models of the speaker and the addressee. With *inderdaad*, the speaker marks a match between his own model of the discourse and her addressee's discourse model: it can be seen as an overt linguistic marker of common ground (“we think alike”). With *eigenlijk*, the speaker marks a contrast between her addressee's discourse model and her own discourse model. Thus, *eigenlijk* can be regarded a linguistic marker of privileged ground (“I know something you don't know”). In addition, *eigenlijk* requires an incorporation of the speech partner's discourse model into one's own discourse model: it expresses that the hearer's expectation is false, but plausible given both speech partners' shared beliefs (“on the basis of our shared knowledge, you couldn't have known what I know”) (see van Bergen et al., 2011, for an extensive analysis). This additional meaning aspect of *eigenlijk* serves a socio-pragmatic goal: the overt acknowledgement of the addressee's perspective on the discourse can be employed as a face saving-strategy in conversational interaction. Whereas speakers generally assume that their conventional implicatures will be uncontroversial (Potts, 2015), note that *inderdaad* and *eigenlijk* express subjective, speaker-based assumptions about the addressee's inferred expectations, which need not be in line with the addressee's actual expectations. Assuming too much or too little about your interlocutor's beliefs can have social consequences in

conversation: for instance, if someone told you *de wereld is eigenlijk rond* ‘the world is actually round’, this message may come across as pejorative.

In sum, *eigenlijk* and *inderdaad* are overt linguistic indices of the speaker's pragmatic inferences about their addressee's linguistic, structural, cognitive and/or social beliefs about the discourse. The central question addressed in this paper is whether and how quickly this inferred content becomes available to the comprehender during incremental language comprehension, which is a hotly debated issue in experimental pragmatics (for reviews, see e.g., Grodner, Klein, Carbary, & Tanenhaus, 2010; Hagoort & Levinson, 2014; Huang & Snedeker, 2018). To our knowledge, this is the first study to investigate effects of such intersubjective markers on online dialogue processing.

#### Discourse markers and discourse processing

Although the number of linguistic analyses of discourse markers has exploded over the last 30 years, relatively little is known about the function of discourse markers in real-time language understanding. In functional approaches, discourse markers are assumed to serve as conversation-managing devices (e.g., Aijmer, 2002; Fischer, 2014; Fox Tree, 2010; Schiffrin, 1987; Smith & Jucker, 2002); that is, they “encode information that is necessary in order to constrain or guide the interpretation process” (Aijmer & Simon-Vandenberg, 2004, pp. 1784). However, considering that language comprehension is a complex, incremental, and combinatorial process, such a theoretical claim lacks empirical evidence as well as specificity with respect to the stage of the interpretation process that would be constrained or guided.

Empirical evidence for their conversation-managing function comes from psycholinguistic investigations of textual discourse connectives, a subset of discourse markers encoding coherence relations between discourse segments. Explicit marking of how two discourse segments should be temporally or causally related has been shown to facilitate processing of subsequent discourse segments (e.g., temporal connectives, Nieuwland, 2015; Politzer-Ahles, Xiang, & Almeida, 2017; causal connectives, e.g., Canestrelli, Mak, & Sanders, 2012; Cozijn, Commandeur, Vonk, & Noordman, 2011; Koornneef & Sanders, 2013; concessive connectives, Xiang & Kuperberg, 2015; and contrastive connectives, Drenhaus, Demberg, Köhne, & Delogu, 2014; Scholman, Rohde, & Demberg, 2017). Other evidence for the facilitating role of discourse markers in online discourse understanding comes from research on focus particles, encoding a relationship between a focused element and its implied alternatives (e.g., Gerwien & Rudka, in press; Kim, Gunlogson, Tanenhaus, & Runner, 2015; Sedivy, 2002).

Whereas the above studies show that some discourse markers can affect online language comprehension beyond the sentence level, none of these studies involved conversational discourse, which is the typical environment of interpersonal DPs in natural language. Linguistic elements that typically occur in conversational interaction and that have been studied experimentally are fillers, such as disfluencies (e.g., *uh*, *uhm*), repairs (e.g., *oh*, *I mean*) and backchannels (e.g., *mhm*, *uhuh*, *really*, *oh*). A range of psycholinguistic studies has shown that fillers can modulate listeners' expectations about, and/or facilitate processing of, upcoming input (e.g., Arnold, Hudson Kam, & Tanenhaus, 2007; Bosker, Quené, Sanders, & de Jong, 2014; Corley, MacGregor, & Donaldson, 2007; Corley, 2010; Fox Tree, 2001; Fox Tree & Schrock, 1999; Lowder & Ferreira, 2016). Various scholars argue that fillers also have a conversation-managing function. For instance, Clark and Fox Tree (2002) have proposed that speakers use disfluencies to announce expected upcoming delays in speech to their addressee. However, this delay-signaling function remains controversial (see e.g., Finlayson & Corley, 2012; Schegloff 2010); others have for instance argued that their facilitating effect on language processing may be better explained as a side effect of speech planning difficulty (e.g., Corley & Stewart, 2008). Regardless of their assumed function, however, there is a crucial difference between disfluencies, repairs and backchannels on the one

hand, and interpersonal DPs on the other. Whether marking a disruption of one's own talk (disfluencies, repairs) or expressing understanding of (generic backchannels) or stance toward (specific backchannels) the other's talk, all these elements encode some *subjective* perspective on the dialogue. By contrast, interpersonal DPs express an *intersubjective* perspective: they encode a relation between speaker beliefs and (assumed) addressee beliefs about the discourse, hence reflecting acknowledgment of mutual belief.

It has been argued by some researchers that language processing is initially egocentric, possibly to minimize processing costs associated with constructing representations of mutual belief (e.g., Horton & Keysar, 1996; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Barr, & Horton, 1998). By contrast, others have demonstrated immediate effects of perspective-taking on language processing in interactive conversation (e.g., Brown-Schmidt & Tanenhaus, 2008; Hanna & Tanenhaus, 2004; Hanna, Tanenhaus, & Trueswell, 2003; for reviews, see Brown-Schmidt & Hanna, 2011; Brown-Schmidt & Heller, 2018). For instance, Brown-Schmidt, Gunlogson, & Tanenhaus (2008) showed in a visual world eye tracking experiment that addressees take the speaker's perspective on the visual context into account (for instance, a grid in which some objects were visible to both speech partners and others are only visible to one interlocutor), and to immediately use this information about common and privileged ground to infer which object the speaker was referring to. Here, we investigate whether listeners are equally sensitive to linguistic common ground-managing cues, i.e., *inderdaad* and *eigenlijk*.

### The present study

We carried out two visual world eye-tracking experiments (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), in which participants were engaged in an online dialogue completion task. They first read an introductory context sentence that served to build up a discourse model ('setting the scene'), and then listened to a short dialogue consisting of a question (or other sequence-initiating action) and a response. Simultaneously, they looked at a visual display with images of four potential discourse referents. The spoken responses in the dialogues either contained one of the two expectation-managing DPs (*inderdaad* and *eigenlijk*) or one of several control adverbs (see Appendix A for examples). Moreover, one word describing a critical discourse referent was left out of the response (replaced by a beep). The task for the participants was to choose the most likely dialogue continuation by clicking on the referent that they thought best fit the position of the beep in the response. We analyzed their click behavior as well as looks to potential dialogue continuations over the time course of the response.

Previous VWP research has convincingly shown that the context strongly constrains potential sentence interpretations online, as measured by a higher probability of fixating contextually likely referents relative to less likely referents (e.g., Tanenhaus et al., 1995) even before the critical referent is mentioned in the speech stream (e.g., Altmann & Kamide, 1999; van Bergen & Flecken, 2017; Kamide, Altmann, & Haywood, 2003; Knoeferle, Crocker, Scheepers, & Pickering, 2005). We aimed to replicate this finding by comparing looks to potential dialogue continuations in strongly constraining dialogues with less constraining dialogues (as measured by offline cloze probabilities) over the time course of the response. More importantly, we built on this finding by examining to what extent expectation-managing DPs modulate the effect of discourse constraint, by comparing looks to contextually likely vs. less likely dialogue continuations in strongly constraining dialogues upon encountering a DP.

We opted for an adapted version of the original visual world paradigm (i.e., clicking on a non-mentioned referent) to do justice to the typical polyfunctionality of expectation-managing DPs. As discussed above, *inderdaad* and *eigenlijk* can apply to various levels of information expressed in their host utterance: they can mark a relation with an

inferred expectation about a discourse referent, but also about a phoneme, an event structure or a speech act. We presented participants with a referential contrast in the visual context, by which we strongly constrained, but still allowed for variation in, potential dialogue interpretations. A more traditional paradigm (i.e., clicking on a mentioned referent) would be able to investigate how participants arrive at a specific referential interpretation of DPs, which itself is an important and highly relevant question. However, our main research question is how comprehenders interpret such inherently polyfunctional cues to begin with, which is as of yet unknown. Moreover, an experimental design that only allows for referential DP interpretations could induce an unnatural interpretation bias, thereby decreasing the ecological validity of the experiments. By investigating click-contingent gaze patterns, we will still be able to assess how listeners arrive at a referential DP interpretation without enforcing this specific interpretation.

We assumed that upon encountering *inderdaad* or *eigenlijk* (relative to a control adverb), listeners would differentially update their beliefs about how the dialogue would continue. As discussed above, *eigenlijk* signals misalignment between what speakers say and what they assume their addressee to expect. From a processing perspective, we hypothesized that *eigenlijk* functions as a warning signal to the comprehender about an upcoming unexpected discourse continuation. In this sense, effects of *eigenlijk* may resemble the effects of contrastive discourse connectives and disfluencies on online language processing: *eigenlijk* is hypothesized to signal the unexpectedness of some aspect of the unfolding dialogue. It is less clear what the effect of *inderdaad* will be on listeners' online dialogue understanding. Upon encountering *inderdaad*, there may be no reason for the listener to change her expectations about the unfolding dialogue, as her discourse model was already established before encountering the linguistic confirmation. Alternatively, listeners could use the linguistic confirmation to strengthen their beliefs about the most likely discourse continuation. This may yield a reduction of cognitive resources spent on distributing probabilities over multiple alternative dialogue continuations during listening.

The basic meanings of *inderdaad* and *eigenlijk* arguably differ in terms of communicative utility. From a processing perspective, a warning signal for any upcoming unexpectedness (*eigenlijk*) is likely more beneficial to the comprehender than a confirmation of an already established expectation (*inderdaad*). Taking frequency of occurrence as a proxy for communicative utility, this difference between *inderdaad* and *eigenlijk* is reflected in natural language: in the *Corpus Gesproken Nederlands* (Corpus of Spoken Dutch; Oostdijk, 2000), a collection of 1000 h of contemporary Dutch speech, *eigenlijk* occurs about 3.5 times more often than *inderdaad*. Taking a subcorpus of approximately 1000 spontaneously produced dyadic conversations between peers, *eigenlijk* occurs about twenty times per 10,000 words, whereas *inderdaad* occurs roughly seven times per 10,000 words. Based on this difference, effects of *inderdaad* on online discourse understanding (if any) are expected to be smaller than effects of *eigenlijk*.

By combining offline and online measures, we investigated effects of interpersonal DPs on discourse interpretations (the selected discourse referent) as well as online discourse processing (i.e., during listening). This allowed us to assess DP effects on distinct stages of the interpretation process, which have not yet been theoretically specified. Similarly, Kurumada, Brown, Bibyk, Pontillo and Tanenhaus (2014) investigated effects of contrastive prosody on reference resolution (*it looks/LOOKS like a zebra*). They found that contrastive pitch elicited early eye-movements to non-prototypical referents (e.g., a zebra-like animal), suggesting that listeners immediately used the prosodic cue and the visual context to generate predictions about the upcoming referent. Despite this immediate effect on visual attention, mouse clicks in the contrastive pitch condition were delayed compared to the non-contrastive pitch condition. Following Grodner et al. (2010), Kurumada et al. (2014) propose that the delayed responses do not reflect the costs of generating a pragmatic inference, but the additional work associated with integrating a more complex discourse interpretation on the basis of

this pragmatic inference with relevant information in the context. In the present study we address this claim: whereas *inderdaad* confirms an already established discourse model, *eigenlijk* requires a revision of an established discourse model. We therefore hypothesize that both expectation-managing DPs immediately modulate visual attention, but that the integration of a revised discourse model with the information in the visual context may incur additional processing costs. This could be reflected in delayed discourse completions after encountering *eigenlijk* compared to encountering *inderdaad*.

## Experiment 1

### Method

#### Participants

Forty native Dutch speakers (31 female, age  $M = 22.4$  years, range 18–33) participated in the first experiment. All were students at Radboud University in Nijmegen, had normal or corrected-to-normal vision and normal auditory acuity. All were paid for participation.

#### Materials and design

**Linguistic stimuli.** Experimental materials comprised 48 short conversations, consisting of a written context sentence followed by a spoken dialogue. Conversations occurred in 4 conditions. The level of discourse constraint was manipulated such that the discourse (the context sentence in combination with the question) either evoked a specific referential expectation (*High*) or not (*Medium*). In high-constraining conversations, the beep in the responses followed either an adverb or adverbial phrase (*High-Adverb*, e.g., *heel erg* ‘very much’, *gisteren* ‘yesterday’) or a discourse particle (*High-Inderdaad* vs. *High-Eigenlijk*); in medium-constraining contexts, responses were identical to the responses in the *High-Adverb* condition (*Medium-Adverb*). An example of an experimental conversation in all conditions is given in Table 1; a subset of 10 stimuli can be found in the Appendix A.

Discourse constraint was determined on the basis of a web-based cloze test, in which participants read each of the 48 experimental conversations in one condition, and were asked to fill in the missing word in the response (12 items per condition, distributed over 4 lists; 20 participants per list). Lemmas with the highest cloze probability in the *High-Adverb* condition (*leeuw* ‘lion’ in the above example) were selected as Target responses; cloze probabilities for these Target responses were then calculated for each condition (Table 2). We also calculated cloze probabilities at the discourse level, whereby all responses that

**Table 1**  
Example of an experimental stimulus.

Constraint	Conversation
High	<p><u>Context</u> <i>Ondanks haar angst voor dieren is Marie naar het circus geweest.</i> Despite her fear of animals, Mary went to the circus.</p> <p><u>Question</u> “<i>Je vond de dierenact zeker doodeng?</i>” You must have been terrified by the animal act?</p> <p><u>Response</u> “<i>Ik schrok heel erg/inderdaad/eigenlijk van de rondrennende *BEEP* aan het eind.</i>” I was <b>very/indeed/actually</b> scared by the running *BEEP* at the end.</p>
Medium	<p><u>Context</u> <i>Op zaterdagmiddag is Marie naar het circus geweest.</i> On Saturday afternoon, Mary went to the circus.</p> <p><u>Question</u> “<i>Wat is je het best bijgebleven?</i>” What did you find most memorable?</p> <p><u>Response</u> “<i>Ik schrok heel erg van de rondrennende *BEEP* aan het eind.</i>” I was <b>very</b> scared by the running *BEEP* at the end.</p>

**Table 2**  
Cloze probabilities of Target responses in each condition.

Condition	Target			
	Lemma		Discourse	
	<i>M</i>	<i>(SD)</i>	<i>M</i>	<i>(SD)</i>
<i>High-Adverb</i>	0.74	(0.22)	0.87	(0.15)
<i>High-Inderdaad</i>	0.78	(0.19)	0.93	(0.10)
<i>High-Eigenlijk</i>	0.52	(0.23)	0.62	(0.24)
<i>Medium-Adverb</i>	0.23	(0.23)	0.28	(0.28)

semantically fit the discourse constraint (e.g., *olifant* ‘elephant’, *tijger* ‘tiger’ in the example above) were counted as Target responses (two independent coders, inter-annotator agreement 97%; disagreements were resolved through discussion; cf. Table 2).

Analyses of participants’ responses on the cloze test were performed by a logistic mixed-effects regression analysis using the *lme4* (Bates, Maechler, Bolker, & Walker, 2015) and *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017) packages in R (R core team, 2015). The final model predicted the probability of a Target response on the basis of Condition (4 levels, *High-Adverb* as reference category), and included the maximal random effects structure justified by the design (Barr, Levy, Scheepers, & Tily, 2013). Results revealed that, compared to *High-Adverb* conversations, Target cloze probabilities were significantly lower in *High-Eigenlijk* conversations (Lemma:  $\beta = -1.41$ ,  $SE = 0.22$ ,  $p < 0.001$ ; Discourse:  $\beta = -2.14$ ,  $SE = 0.28$ ,  $p < 0.001$ ) and *Medium-Adverb* conversations (Lemma:  $\beta = -3.41$ ,  $SE = 0.33$ ,  $p < 0.001$ ; Discourse:  $\beta = -3.94$ ,  $SE = 0.41$ ,  $p < 0.001$ ). Target cloze probability was significantly higher in *High-Inderdaad* conversations than in *High-Adverb* conversations at the Discourse level ( $\beta = 0.77$ ,  $SE = 0.37$ ,  $p = 0.04$ ); the difference did not reach significance at the Lemma level ( $\beta = 0.18$ ,  $SE = 0.18$ ,  $p = 0.3$ ). These results suggest that both *inderdaad* and *eigenlijk* influenced the likelihood that the dialogue would continue with the target word (e.g., *leeuw* ‘lion’), compared to the same dialogue without a DP (but with an adverb). Note, however, that Target cloze probabilities were still relatively high in the *High-Eigenlijk* condition, suggesting that *eigenlijk* was not always interpreted as marking a referential contrast. These results underline the highly context-dependent nature of conventional implicatures: the inferred expectation that *eigenlijk* contrasts with can be based on any aspect of the surrounding discourse situation.

Spoken dialogues were recorded by four Dutch native speakers (two male, two female), in all possible questioner-answerer combinations. Because the DPs typically appear in casual conversation, and not in formal or read speech, care was taken to mimic natural dialogue as closely as possible. To this end, conversational pairs sat in a recording booth together while their interactions were being recorded. They were instructed not to read the utterances from paper, but to speak informally and to imagine having a real conversation. Responses were recorded as complete utterances (i.e., including the target word) to keep the sentence intonation pattern as natural as possible; to avoid co-articulation, speakers were told to pause before pronouncing the critical word. Each question-answer pair was recorded three times, the most natural ones of which were selected as experimental materials. All sound files were normalized in intensity; critical words in the responses were replaced by a 440 Hz sine waveform with a length of 500 ms using version 2.1.1 of Audacity®.

**Visual stimuli.** 48 visual displays were constructed, each containing four images that appeared randomly in one of the four corners of a 22-in. computer screen. Each display contained one referent that was highly expected on the basis of the discourse (Expected), one referent that was contextually unexpected but related to the discourse (Unexpected) and two referents that were unrelated to the discourse (Distracters).

Expected referents depicted the lemma with the highest cloze probability in the *High-Adverb* condition (*leeuw* 'lion' in the above example). Unexpected referents were chosen from the discourse completions that were incongruent with the contextually most likely dialogue continuation in the *High-Adverb* Condition. Note that there was considerable variation in Unexpected dialogue completions: for instance, if the contextually most likely referent was a specific type of food (e.g., *hamburger*), any other type of food would qualify as an unexpected referent (e.g., *pizza*, *pasta*, *sushi*, *ice cream*, ...). Moreover, depictability was an important selection criterion, as well as the additional constraint that each referent could occur only once in the experiment. Finally, 'Unexpected' referents should be roughly equally predictable as 'Expected' referents in medium-constraining dialogues.

Because of these constraints, Unexpected referents were not always the lemma with the highest cloze probability in the *High-Eigenlijk* condition. Still, relative to the *High-Adverb* condition (Lemma:  $M = 0.03$ ,  $SD = 0.16$ ; Discourse:  $M = 0.09$ ,  $SD = 0.13$ ), Unexpected referents had a higher cloze probability in the *High-Eigenlijk* condition (Lemma:  $M = 0.10$ ,  $SD = 0.15$ ,  $\beta = 1.65$ ,  $SE = 0.25$ ,  $p < 0.001$ ; Discourse:  $M = 0.35$ ,  $SD = 0.24$ ,  $\beta = 2.05$ ,  $SE = 0.15$ ,  $p < 0.001$ ), and a lower cloze probability in the *High-Inderdaad* condition (Lemma:  $M = 0.004$ ,  $SD = 0.03$ ,  $\beta = -1.99$ ,  $SE = 0.55$ ,  $p < 0.001$ ; Discourse:  $M = 0.03$ ,  $SD = 0.05$ ,  $\beta = -1.28$ ,  $SE = 0.23$ ,  $p < 0.001$ ). In the *Medium-Adverb* condition, Lemma cloze probabilities did not significantly differ between Unexpected ( $M = 0.18$ ,  $SD = 0.20$ ) and Expected referents,  $\beta = 0.24$ ,  $SE = 0.37$ ,  $p = 0.53$  (Discourse cloze values were not calculated for unexpected referents, as there was no clear discourse constraint to contrast these referents with). Distracter referents had a cloze probability of zero in every condition. An example of a display is given in Fig. 1.

Photographs were either self-taken or acquired from the Bank of Standardized stimuli (BOSS; Brodeur, Dionne-Dosti, Montreuil, & Lepage, 2010), the photo-stock website Dreamstime.com, or through the Google Image search engine. Photos were edited to reduce visual complexity (e.g., backgrounds removed where possible), converted to grayscale and resized to a square format of  $250 \times 250$  pixels.



Fig. 1. Example of a display. NB: Labels (for illustration purposes only) refer to the expectation evoked by the discourse, not by the DP in the response.

*Filler items.* In addition to the experimental items, 48 filler dialogues were created, following the same structure as the experimental dialogues. Filler discourses varied from low- to high-constraining; responses never contained a DP. The filler items were combined with the same displays as the experimental conversations, in which any picture could serve as Expected, Unexpected or Distracter referent.

*Design.* Experimental items were counterbalanced across four lists following a Latin square design, such that each participant encountered each experimental conversation in only one condition (12 items per condition). The 48 filler items were added to each list, yielding a total of 96 items per list. Trial order was pseudo-randomized, such that each display occurred once in the first half of the experiment and once in the second half. Half of the participants first encountered the display combined with the experimental conversation; the other half first saw the display in combination with the filler conversation.

#### Procedure

Participants were tested at the Max Planck Institute for Psycholinguistics in Nijmegen. All gave written consent to participate in the experiment. The study was ethically approved by the Faculty of Social Sciences at Radboud University Nijmegen (Ethics Approval #ECG2013-1308-120).

To ensure that participants would correctly identify the photographs used in the experiment, they first performed a picture familiarization task in which all photos were presented one-by-one, accompanied by a written one-word description. This task was self-paced, and took on average 5 min.

Eye movements of the dominant eye were recorded using a tower-mounted eye-tracker with a 1000 Hz sampling rate (Eyelink 1000; SR Research). Participants sat about 100 cm from the monitor with their heads in a chin rest; a 9-point calibration was conducted for each participant at the beginning of the experiment. Auditory stimuli were played via headphones, which participants were instructed to wear throughout.

Each trial began with a fixation cross that remained on the screen for 500 ms, after which the written context sentence was presented. Once they had read the context, participants pressed a button; 500 ms later the question was played. 300 ms after question offset the display appeared on the screen; after a preview of 1500 ms, the response was played; simultaneously, a cursor appeared in the center of the display. Participants were instructed to click on the picture in the display that they thought best fit at the position of the beep in the response. They could only click on a picture after the response had been fully played. After one of the pictures was clicked, the display disappeared and the next trial started.

The experiment consisted of four practice trials, followed by four blocks of 24 experimental trials; participants took self-timed breaks in between. Each experimental session took approximately 45 min in total.

#### Results

We first assessed whether and how comprehenders' understanding of the conversation was affected by the discourse constraint and the discourse particles by analyzing dialogue completion responses (selected referents). Trials in which a Distracter referent was selected were excluded from further analyses (0.3% of the data; analyses on the full data set yielded similar results); excessive RTs ( $\pm 2.5 SD$ ) were also removed (2.8% data loss). One experimental item had to be excluded because the question-answer pair was accidentally combined with the wrong (i.e., medium-constraining) context sentence in the *High-Inderdaad* condition, resulting in 1% data loss. Data analyses were performed by means of (generalized) linear mixed effects regression models using the *lme4* and *lmerTest* package in R (R core team, 2015).

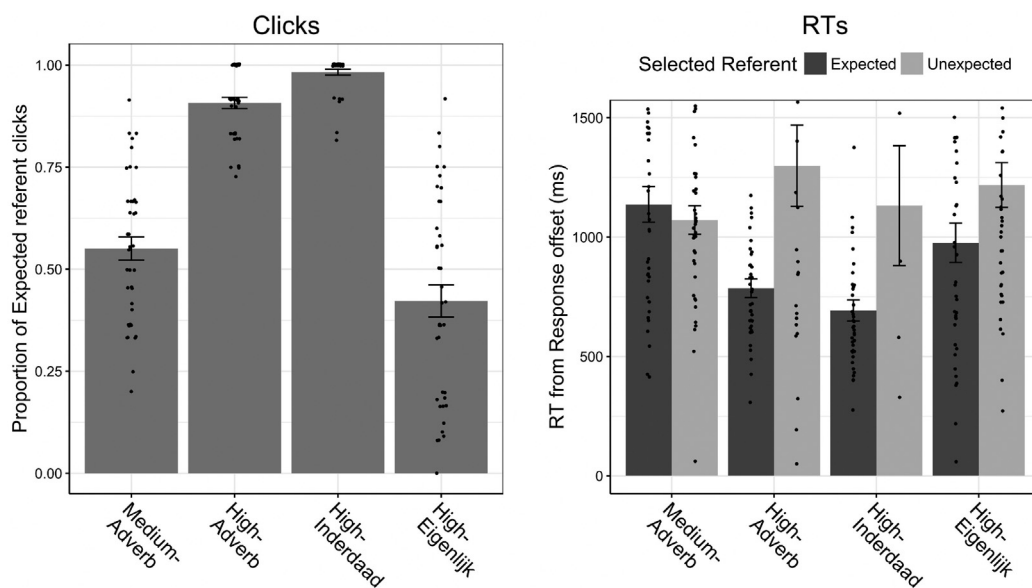


Fig. 2. Proportion of Expected vs. Unexpected referent choices (left panel) and response latencies (right panel) in each condition. Error bars represent  $\pm$  SE.

### Behavioral results

Fig. 2 shows the proportion of Expected (as opposed to Unexpected) referent clicks in each condition (left panel) and the mean RTs from Response Offset per condition split up by selected referent (right panel). Jittered dots represent mean proportions/RTs per participant.

The left panel suggests an effect of contextual constraint on the referential choice: the preference to select the Expected referent is increased in the *High-Adverb* condition compared to the *Medium-Adverb* condition (where there is no contextual expectation). This preference is even stronger in *High-Inderdaad* contexts, whereas in the *High-Eigenlijk* condition there is an increased preference to click on the Unexpected referent. Results from a logistic mixed-effects analysis<sup>1</sup> including Condition (4 levels, reference category *High-Adverb*) as fixed predictor confirm that all condition levels significantly differ from the *High-Adverb* condition: the probability of Expected referent clicks is significantly higher in the *High-Inderdaad* condition ( $\beta = 1.94$ ,  $SE = 0.40$ ,  $p < 0.001$ ), and significantly lower in the *High-Eigenlijk* condition ( $\beta = -3.09$ ,  $SE = 0.22$ ,  $p < 0.001$ ) as well as the *Medium-Adverb* condition ( $\beta = -2.47$ ,  $SE = 0.21$ ,  $p < 0.001$ ).

The right panel suggests an effect of contextual constraint on response latencies as well: for high-constraining contexts, unexpected dialogue completions take more time than expected dialogue completions, whereas there seems to be no latency difference in medium-constraining contexts. Note that the number of unexpected discourse completions is very small in the *High-Adverb* and *High-Inderdaad* condition, because dialogues in these conditions were almost always completed with the expected referent. When comparing expected discourse completion latencies between the three high-constraining conditions, responses seem slowest in the *High-Eigenlijk* condition, suggesting that encountering *eigenlijk* delays the referential choice. Linear mixed-effects analyses predicting (log-transformed) reaction times on the basis of Condition, Selected Referent (Expected vs. Unexpected) and their interaction revealed a main effect of Selected Referent ( $\beta = -0.40$ ,  $SE = 0.09$ ,  $p < 0.001$ ), indicating that participants were generally faster to complete dialogues with Expected referents than with Unexpected referents. The interaction effect was also significant (comparing models with vs. without interaction effect:  $X^2(3) = 12.93$ ,

$p < 0.001$ ), indicating that the effect of Selected Referent differed between conditions. Follow-up analyses revealed no significant differences between conditions for Unexpected discourse completions (all  $p$ 's  $> 0.25$ ); for Expected discourse completions, reaction times were shorter in the *High-Inderdaad* condition compared to the *High-Adverb* condition ( $\beta = -0.26$ ,  $SE = 0.07$ ,  $p < 0.001$ ). There was no evidence for a latency difference between the *High-Eigenlijk* condition and the *High-Adverb* condition ( $\beta = 0.12$ ,  $SE = 0.08$ ,  $p = 0.15$ ), but expected discourse completions were slower after encountering *eigenlijk* than after encountering *inderdaad* ( $\beta = 0.38$ ,  $SE = 0.09$ ,  $p < 0.001$ ).

### Interim summary

Findings from the discourse completion task confirm that expectation-managing DPs affect listeners' beliefs about dialogue continuations. Encountering *inderdaad*, confirming an inferred expectation, led to more Expected dialogue completions; encountering *eigenlijk*, marking a contrast with an inferred expectation, resulted in more Unexpected dialogue completions.

Our stimuli were designed as to make a referentially contrastive interpretation of *eigenlijk* most prominent (i.e., a contrastive referent was consistently present in the visual context; higher cloze probability for Unexpected referents in *High-Eigenlijk* stimuli), but *eigenlijk* was only interpreted as marking a referential contrast in around half of the cases. This means that, despite its strong constraint on possible interpretations, the visual context did not impose a referential interpretation of *eigenlijk* on the listener.

Moreover, note that we found considerable variation between participants in the discourse completion responses in the *High-Eigenlijk* condition. Some participants were strongly biased to interpret *eigenlijk* referentially, almost exclusively choosing Unexpected discourse completions, whereas others hardly ever selected the Unexpected referent. At the same time, there was considerable item variation, some stimuli eliciting many more Expected dialogue completions than others. For instance, an experimental dialogue that almost exclusively elicited Expected clicks in the *High-Eigenlijk* condition is given in (2):

- (2) **Context**  
*Vlak voor zijn sollicitatiegesprek heeft Maarten een broodje met pesto gegeten.*  
 Right before his job interview, Maarten ate a sandwich with pesto.  
**Question**

<sup>1</sup> The model including the maximal random effects structure failed to converge; reported estimates are based on the final model including random intercepts for participants and items.

Heb ik nog iets tussen mijn tanden?

Is there something between my teeth?

Response

*Ik zie eigenlijk daar rechts iets in je (mond /haar) zitten.*

I see [actually] something there on the right in your (mouth/ hair).

In this case, *eigenlijk* is likely interpreted as signaling a socio-pragmatic contrast; that is, a contrast between the speech act (an affirmative response) and an implied expectation for a socially preferred no-response. In principle, the same variability holds for the interpretation of *inderdaad*, i.e., relating to *some* discourse aspect but not necessarily a referential aspect. Unlike *eigenlijk*, however, *inderdaad* does not mark a *contrastive* relation. Rather, it reduces the probability of a contrast with any discourse aspect, including a referential contrast. The choice for an Unexpected dialogue completion in the *High-Inderdaad* condition would therefore be highly unlikely.

Whereas *inderdaad* was found to speed up Expected discourse completions relative to adverbs, we found no evidence for a speed-up effect of *eigenlijk* if listeners committed to a referential interpretation. In fact, Unexpected dialogue completions were slower than Expected completions even after encountering *eigenlijk*. This could mean that participants entertained the two referentially contrastive interpretations all along and made up their mind only shortly before selecting a picture. Moreover, we found no evidence that response latencies differed between *eigenlijk* and adverbs, leaving open the possibility that participants ignored *eigenlijk* in these cases. In order to investigate whether and how listeners integrated the expectation-managing information encoded in DPs during the online interpretation process, we examined listener's gaze behavior as the conversation unfolded.

### Click-contingent gaze patterns

Fixations to the four areas of interest (AoIs), i.e., the four referents in the display, were calculated for the duration of the Response, using the EyeLink Data Viewer software (version 2.1.1, SR Research). We analyzed fixations in relation to listeners' click behavior. Analyzing click-contingent gaze patterns allows for an assessment of how listeners arrived at their final interpretation, and to what extent DPs influenced this interpretation process.

Fig. 3 shows the proportion of click-contingent AoI fixations from Response onset until Response offset per condition; grey-shaded areas indicate  $\pm$  SEs. Solid lines represent the proportion of looks to the eventually clicked referent; dashed lines show the proportion of looks to the competitor. Dotted lines represent the average proportion of looks to the two Distracters. Note that the selected referent in the *High-Inderdaad* condition was almost exclusively the Expected referent (e.g., lion). By contrast, in the *High-Eigenlijk* condition, the selected referent was sometimes the Expected and sometimes the Unexpected referent (e.g., clown).

We analyzed two subsequent time windows within the Response time window:

- An *EARLY INTEGRATION* time window, ranging from 200 ms after DP (or adverb) onset until 200 ms after the onset of the beep (mean duration 1191 ms). In this time window we assessed immediate effects of contextual constraint and DPs on visual attention. If listeners immediately integrate the expectation-managing information encoded in the DPs, fixations to the two discourse-related referents are predicted to diverge between conditions in this time window, such that eventually selected referents are fixated more after

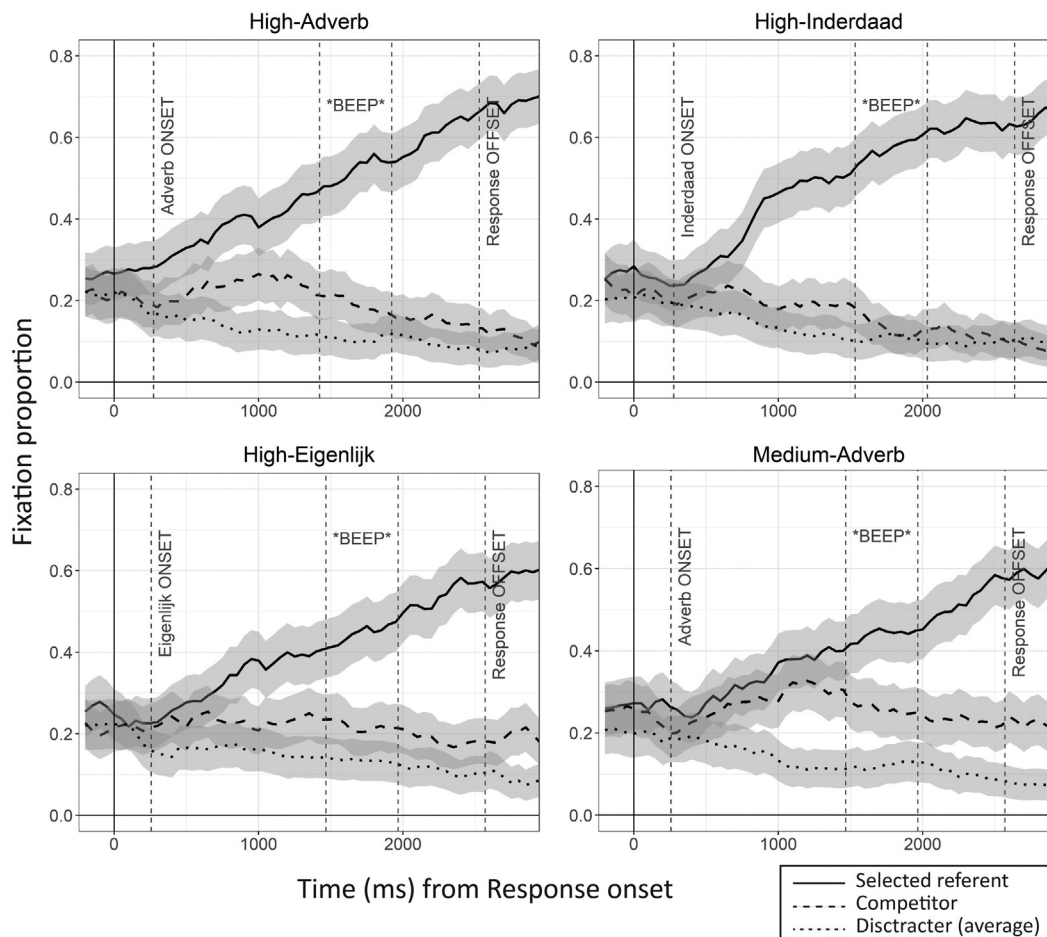
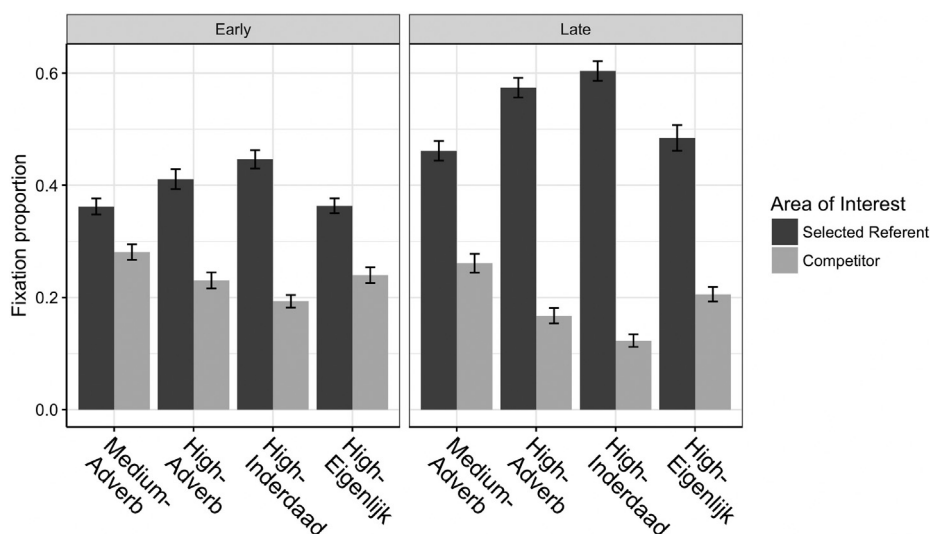


Fig. 3. Click-contingent gaze patterns over time for the duration of the Response.



**Fig. 4.** Proportions of looks to the clicked referent (dark grey bars) and the competitor (light grey bars) per condition during the Early (left panel) and Late Integration window (right panel).

encountering *inderdaad* and less after encountering *eigenlijk*.

- A *LATE INTEGRATION* time window, ranging from 200 ms after beep onset until Response offset (mean duration 1104 ms); i.e., just before listeners select the referent best fitting the Response. AoI fixations during this part of the Response are analyzed to assess later and/or sustained effects of discourse constraint and DPs on listeners' final discourse interpretation. If integration of the DPs requires additional processing, fixations to the two discourse-related referents are predicted to start diverging between conditions in this time window.

For statistical analysis, we calculated the (log-transformed) odds of fixating the Selected referent over the Competitor by subtracting the odds of fixating the Competitor referent from the odds of fixating the Selected referent (cf. Rommers, Meyer, Praamstra, & Huettig, 2013) in each time window. Linear mixed effects regression analyses were performed to predict the probability of fixating the eventually Selected referent over the Competitor on the basis of *Condition*. Unless indicated otherwise, final models included the maximal random effects structure justified by the design (cf. Barr et al., 2013).

Fig. 4 shows fixation proportions of Selected Referents and Competitors in each condition for the two time windows. The figure suggests a general preference to fixate the eventually selected referent; this preference seems to differ between conditions, suggesting an effect of both contextual constraint and expectation-managing DPs on the interpretation process. The differences between conditions seem to exist in both time windows, but appear to be more pronounced in the Late Integration window.

We first analyzed the effect of contextual constraint by means of a linear mixed-effects regression model predicting the probability of Clicked Referent fixations over Competitor fixations on the basis of *Constraint* (2 levels; *Medium-Adverb* (intercept) vs. *High-Adverb*) in the two time windows. In the Early Integration window, results showed a preference to fixate the to-be-clicked referent over the competitor in the *Medium-Adverb* condition ( $\beta = 1.15$ ,  $SE = 0.34$ ,  $p < 0.01$ ); this preference was stronger in the *High-Adverb* condition ( $\beta = 1.21$ ,  $SE = 0.49$ ,  $p < 0.05$ ). In the Late Integration window, the difference between medium and high-constraining contexts was also significant ( $\beta = 2.87$ ,  $SE = 0.56$ ,  $p < 0.001$ ). These results suggest that the prior context constrains visual attention: attentional preference to the eventually selected referent is increased if the context evokes a specific referential expectation.

Next, we analyzed effects of expectation-managing DPs on visual

attention in high-constraining contexts by analyzing the probability of Clicked Referent over Competitor fixations on the basis of *Condition* (3 levels) in the two time windows. Results show that the difference between *Inderdaad* and *Eigenlijk* was significant in both time windows (*Early*:  $\beta = 1.48$ ,  $SE = 0.47$ ,  $p < 0.01$ ; *Late*:  $\beta = 3.00$ ,  $SE = 0.67$ ,  $p < 0.001$ ), indicating that the preference to fixate the to-be-clicked referent is stronger after encountering *inderdaad* than after encountering *eigenlijk*. We found no evidence for a difference between either of the DP conditions and the *High-Adverb* condition in the Early Integration window (*Inderdaad*:  $\beta = 0.77$ ,  $SE = 0.42$ ,  $p = 0.07$ ; *Eigenlijk*:  $\beta = -0.48$ ,  $SE = 0.48$ ,  $p = 0.15$ ); in the Late Integration window, the differences did reach significance (*Inderdaad*:  $\beta = 1.04$ ,  $SE = 0.49$ ,  $p < 0.05$ ; *Eigenlijk*:  $\beta = -1.96$ ,  $SE = 0.65$ ,  $p < 0.01$ ), suggesting that both DPs affect visual attention to potential discourse continuations.

Recall that for conversations in the *High-Eigenlijk* condition, the choice between Expected and Unexpected referents was more or less equally distributed; arguably, both could be likely dialogue completions in this condition, depending on the inferred expectation that *eigenlijk* was interpreted to mark a contrast with. In the *High-Adverb* and the *High-Inderdaad* condition, however, participants predominantly selected the referent that was most likely given the context (Expected referent), leaving us with too few data points to analyze Unexpected click-contingent gaze behavior in these conditions.

In order to assess to what extent gaze patterns are dependent on the referent choice, we investigated gaze patterns for Expected and Unexpected referent clicks separately for the *High-Eigenlijk* condition (Fig. 5). Note that the solid lines in the two panels now refer to different referents, i.e., the Expected referent in the left panel and the Unexpected referent in the right panel.

There seems to be an initial preference to fixate the Expected referent (solid line in left panel; dashed line in right panel), irrespective of the eventual referent choice, in the first 500 ms after Response onset, suggesting an effect of contextual constraint on visual attention. If the Unexpected referent is eventually clicked (right panel), attention to the Unexpected referent starts increasing very quickly after encountering *eigenlijk*. This suggests that the initial contextual expectation is revised immediately upon encountering the DP. This is corroborated by an analysis predicting the probability of fixating the Selected referent over the Competitor on the basis of *Referent Choice* (Expected vs. Unexpected, centered), showing an early preference to fixate the to-be-clicked referent over the competitor ( $\beta = 1.57$ ,  $SE = 0.32$ ,  $p < 0.001$ ); this preference does not differ between Expected and Unexpected



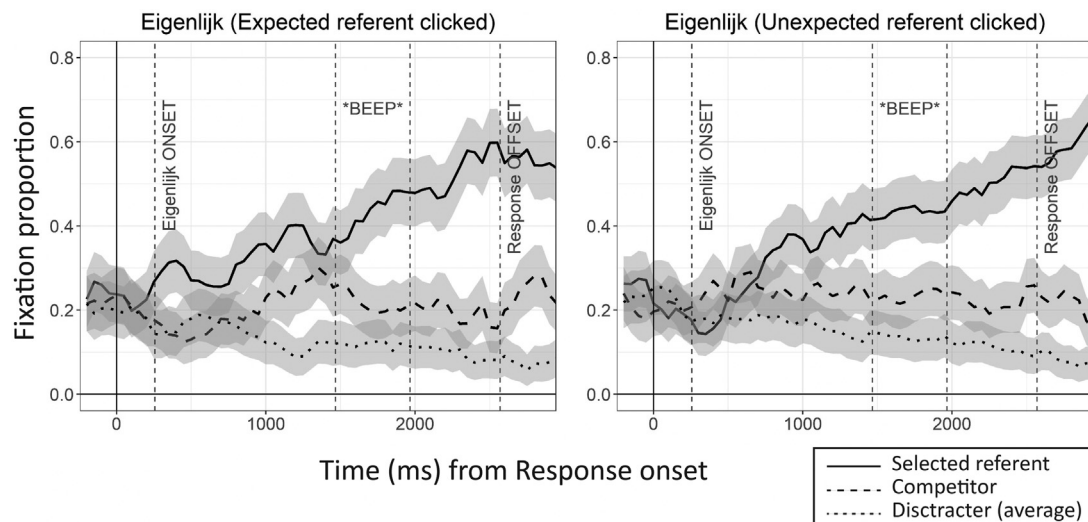


Fig. 5. Click-contingent gaze patterns over time for Expected (left) and Unexpected (right) referent clicks in the High-Eigenlijk condition. NB: solid lines refer to Expected referents in the left panel, and to Unexpected referents in the right panel.

referents ( $\beta = 0.56$ ,  $SE = 0.70$ ,  $p = 0.43$ ), indicating that the initial attentional preference is immediately revised when *eigenlijk* is interpreted referentially. Moreover, both panels in Fig. 5 suggest sustained attention to the semantic competitor throughout the Response time window (when compared with the *High-Adverb* and *High-Inderdaad* condition, cf. Fig. 3). This suggests that competition from alternative dialogue continuations is stronger after encountering *eigenlijk* than after hearing *inderdaad* or an adverb, regardless of its interpretation.

Restricting our analysis of the three high-constraining conditions to trials in which the Expected referent was chosen, results show that the preference to fixate the Expected referent is stronger after *inderdaad* than after *eigenlijk* in both time windows (*Early*:  $\beta = 1.32$ ,  $SE = 0.56$ ,  $p < 0.05$ ; *Late*:  $\beta = 3.01$ ,  $SE = 0.64$ ,  $p < 0.001$ ). We found no evidence for a difference between either DP condition and the *Adverb* condition in the Early Integration window (*Inderdaad*:  $\beta = 0.63$ ,  $SE = 0.40$ ,  $p = 0.12$ ; *Eigenlijk*:  $\beta = -0.70$ ,  $SE = 0.55$ ,  $p = 0.21$ ); we can hence not determine whether the early difference between the DPs is due to increased competition from alternative referents after encountering *eigenlijk*, or reduced competition from alternative referents after hearing *inderdaad*. In the Late Integration window, however, both DPs significantly differ from adverbs (*Inderdaad*:  $\beta = 0.96$ ,  $SE = 0.47$ ,  $p < 0.05$ ; *Eigenlijk*:  $\beta = -2.05$ ,  $SE = 0.67$ ,  $p < 0.01$ ), confirming that both *inderdaad* and *eigenlijk* modulate attentional preferences for the Expected referents when compared with adverbs.

To assess effects of referentially contrastive interpretations of *eigenlijk* on gaze behavior, we compared Unexpected click-contingent gaze patterns in the *High-Eigenlijk* condition with Expected click-contingent gaze patterns in the *High-Inderdaad* and *High-Adverb* condition. Analyses of both time windows reveal that the preference to fixate the to-be-clicked Unexpected referent after encountering *eigenlijk* is weaker than to fixate the to-be-clicked Expected referent after encountering *inderdaad* (*Early*:  $\beta = 1.79$ ,  $SE = 0.49$ ,  $p < 0.001$ ;<sup>2</sup> *Late*:  $\beta = 3.05$ ,  $SE = 0.85$ ,  $p < 0.001$ ) or an adverb (*Early*:  $\beta = -1.15$ ,  $SE = 0.58$ ,  $p = 0.05$ ); *Late*:  $\beta = -1.99$ ,  $SE = 0.72$ ,  $p < 0.01$ ). This suggests that if listeners commit to a referential interpretation of *eigenlijk*, they still experience increased competition from alternative dialogue continuations when compared with an adverb or *inderdaad*.

In sum, findings from the click-congruent gaze patterns provide evidence that both *eigenlijk* and *inderdaad* affect visual attention to potential dialogue continuations during listening, such that *inderdaad*

reduces and *eigenlijk* increases competition from alternative dialogue continuations when compared with an adverb. We found no evidence for an immediate effect of *inderdaad* on visual attention when compared with adverbs; by contrast, *eigenlijk* immediately increased attention to competing discourse referents relative to adverbs, but only if comprehenders committed to a referentially contrastive dialogue continuation.

#### Discussion

Combining the gaze patterns with the behavioral results, findings confirm our hypothesis that *inderdaad* reduces competition from alternative interpretations during listening (although not immediately in the Early Integration window), suggesting that the inferred expectation confirmed by *inderdaad* was readily available to strengthen listeners' beliefs about possible dialogue continuations, which speeded up Expected dialogue completions relative to adverbs.

Although we found no evidence for effects of *eigenlijk* relative to adverbs on dialogue completion latencies, the gaze patterns confirm that listeners do integrate the contrastive information encoded in *eigenlijk* to guide their visual attention. If participants chose for an Expected dialogue completion after encountering *eigenlijk*, we found increased competition from alternative interpretations relative to adverbs, arguing against participants ignoring *eigenlijk* in these cases. In addition, gaze patterns revealed that participants immediately committed to referentially contrastive interpretations of *eigenlijk*, excluding the possibility that participants entertained the two possible interpretations all along. This did however not result in faster discourse completions: rather, referential interpretations of *eigenlijk* slowed down discourse completions when compared with non-referential interpretations. With this finding we corroborate earlier claims that pragmatic inferences can be generated immediately, whereas the integration of a revised discourse model with information in the visual context incurs additional processing costs (e.g., Grodner et al., 2010; Kurumada et al., 2014).

The considerable item variation in dialogue completions suggests that, despite the prominence of a referential constraint, specific interpretations of *eigenlijk* are determined by the specific characteristics of each individual dialogue, providing empirical evidence for their assumed polyfunctionality and the context-dependent nature of conventional implicatures. Moreover, the large inter-subject variation implies that interpretations of *eigenlijk* are not only situation-specific, but also listener-specific: some participants were much more likely to select Unexpected discourse continuations than others (and vice versa). This

<sup>2</sup> The Early Integration model included random intercepts for participants and items, as well as random by-item slopes for *Condition*.

raises the question where such individual interpretational biases come from. One possibility is that listener-specific interpretation biases may relate to the individual frequency distribution of referential and non-referential uses of *eigenlijk* encountered in everyday conversational interaction. For instance, if an individual has encountered mainly socio-pragmatic (i.e., face-saving) uses of *eigenlijk*, this specific distribution may affect her preferred interpretation of *eigenlijk* in any context. In fact, our experimental design may have also induced an interpretational bias: although dialogue interpretations were left open, stimuli were constructed as to make a referential interpretation most prominent. As a result, some participants may have more strongly favored a referential interpretation of *eigenlijk*.

Another explanation for the inter-subject variation may be that the interpretation of *eigenlijk* is related to individuals' pragmatic competence. Recall that using interpersonal DPs entails perspective-taking: the speaker has to make an inference about her addressee's expectations of the discourse situation. Similarly, when listening to the experimental dialogues, participants need to infer the questioner's expectation that *eigenlijk* in the response contrasts with. Previous research has demonstrated that online pragmatic language comprehension is modulated by individuals' pragmatic abilities (e.g., counterfactuals, Kulakova & Nieuwland, 2016; scalar implicatures, Nieuwland, Ditman & Kuperberg, 2010). If some participants in our sample were more pragmatically competent than others, this may have induced variation in interpretations of *eigenlijk*.

Experiment 2 was designed to further investigate the listener-specific variation in interpretations of *eigenlijk*. We examined (1) the potential role of the frequency distribution of uses of *eigenlijk* in the local experimental environment and (2) the contribution of individual participants' pragmatic competence to their interpretation of *eigenlijk*. To address the first question, we increased the frequency of non-referential uses of *eigenlijk* throughout the experiment, thus allowing us to investigate to what extent a change in the frequency distribution of uses of *eigenlijk* in the experimental environment affects participants' interpretations of *eigenlijk* in high-constraining conversations.

Two hypotheses were formulated. Under the assumption that listeners adapt to the statistics of the experimental environment during language comprehension (e.g., Delaney-Busch, Morgan, Lau, & Kuperberg, 2018; Fine, Jaeger, Farmer, & Qian, 2013, but see Harrington Stack, James, & Watson, 2018; Kaschak & Glenberg, 2004; Norris, McQueen, & Cutler, 2003), we may expect an increased number of non-referential interpretations in the *High-Eigenlijk* condition (i.e., more Expected discourse continuations) in Experiment 2 compared to Experiment 1. During listening, we would expect the effect of *eigenlijk* on visual attention to be immediate, but potentially less pronounced: the increased frequency of non-referential uses of *eigenlijk* may reduce semantic competition from alternative discourse referents, as reflected in decreased attention to Unexpected referents and faster discourse continuations compared to Experiment 1. Alternatively, encountering more contrastive cues in the speech stream altogether (i.e., higher frequency of *eigenlijk* throughout an experimental session) could increase the salience of the referential contrast present in the visual context. This may yield a higher proportion of referential interpretations of *eigenlijk*; i.e., more Unexpected discourse continuations compared to Experiment 1. If this is the case, we expect the effect of *eigenlijk* on visual attention to again be immediate, but potentially more pronounced: we may find increased attention to semantic competitors during listening and delayed discourse continuations compared to Experiment 1.

To assess whether the interpretation of interpersonal DPs is related to individuals' pragmatic competence, participants in Experiment 2 were asked to complete two subscales of the Autism-Spectrum Quotient (AQ) questionnaire, a self-assessment screening instrument to assess the degree to which an individual of normal intelligence shows autistic traits (Baron-Cohen, Wheelwright, Skinner, & Martin, 2001; Dutch translation validated by Hoekstra, Bartels, Cath, & Boomsma, 2008).

This measure has been employed previously in online language tasks to assess individuals' pragmatic language processing. Scores on this scale have for instance been found to modulate the N400 component in ERP studies (Kulakova & Nieuwland, 2016; Nieuwland et al., 2010). The two selected subscales assess social skill and communication, two areas relevant to pragmatic competence. This will allow us to correlate participants' individual pragmatic competence with effects of interpersonal DPs on discourse understanding. Participants with reduced pragmatic skills may overall be less sensitive to interpersonal DPs during language comprehension, in the sense that they are less likely to make pragmatic inferences about the questioner's expectation that *eigenlijk* contrasts with when listening to the dialogue. This could lead to a smaller proportion of Unexpected discourse continuations, and smaller effects of *eigenlijk* on visual attention and response latencies. Pragmatic inferencing could also be cognitively more costly for participants with reduced pragmatic abilities, which could be reflected in delayed effects of *eigenlijk* on visual attention. Alternatively, participants with reduced pragmatic competence may be less sensitive to specifically socio-pragmatic (face-saving) uses of *eigenlijk*. This could be reflected in a higher preference for referential interpretations of *eigenlijk*, as well as more pronounced effects of *eigenlijk* on visual attention and response latencies (i.e., increased attention to alternative discourse referents and delayed response latencies).

## Experiment 2

In Experiment 2 we increased the frequency of non-referential uses of *eigenlijk* by combining medium-constraining contexts with responses containing *eigenlijk* (hereafter: the *Medium-Eigenlijk* condition), as shown in Table 3.

Recall that for medium-constraining dialogues, both 'Expected' (lion) and 'Unexpected' (clown) referents are likely discourse continuations. Since the context does not evoke an expectation of one referent over the other, *eigenlijk* cannot mark a referential contrast in such contexts. Yet, it can express a contrastive relation with some other inferred expectation: in the above example (which is grammatical and pragmatically acceptable in Dutch), it could for instance mark the unexpectedness of the response as a whole, i.e., the responder admitting to be frightened in the circus. Given that there is no specific referential expectation in medium-constraining contexts, we do not expect the presence of *eigenlijk* to affect the referential choice in this condition compared to the *Medium-Adverb* condition in Experiment 1. However, the contrastive meaning of *eigenlijk* may affect visual attention in this condition: we may find the attentional preference for the to-be-selected referent to be delayed compared to the *Medium-Adverb* condition in Experiment 1.

## Method

### Participants

Thirty-two native Dutch speakers (26 female, age  $M = 22.7$  years,

**Table 3**

Example of an experimental stimulus in the *Medium-Eigenlijk* condition in Experiment 2.

Constraint	Conversation
Medium	<p><u>Context</u> Op zaterdagmiddag is Marie naar het circus geweest. On Saturday afternoon, Mary went to the circus.</p> <p><u>Question</u> "Wat is je het best bijgebleven?" What did you find most memorable?</p> <p><u>Response</u> "Ik schrok <i>eigenlijk</i> van de rondrennende *BEEP* aan het eind." I was <b>actually</b> scared by the running *BEEP* at the end.</p>

range 19–33) participated in Experiment 2. All were students at Radboud University in Nijmegen, had normal or corrected-to-normal vision, normal auditory acuity, and were paid for participation.

### Materials and design

We used the same linguistic stimuli as in Experiment 1, except that responses in the *Medium-Eigenlijk* condition were now identical to the responses in the *High-Eigenlijk* condition (but combined with a less constraining context; see Table 3 above). Visual stimuli and filler items were identical to those in Experiment 1; the design and procedure were also identical, except that participants in Experiment 2 completed the shortened version of the AQ questionnaire directly after the experiment. This questionnaire comprised 20 propositions addressing social skill (e.g., *I find it difficult to work out people's intentions*) and communication (e.g., *Other people frequently tell me that what I've said is impolite, even though I think it is polite*). Participants had to indicate on a 4-point scale to what extent they agreed with each proposition (1 = completely agree, 4 = completely disagree). Completing the questionnaire took an additional 2 min on average.

### Results

Data were analyzed using the same analysis procedures as in Experiment 1. Moreover, to compare the results between experiments, we performed additional analyses on the combined data from Experiments 1 and 2 that included the factor *Experiment* and its interaction with other independent variables as fixed predictors.

AQ scores were calculated following the procedure described in Baron-Cohen, Wheelwright, Skinner, Martin, and Clucley (2001); higher AQ scores indicate stronger autistic traits. For each AQ subscale the maximal score is 10; participants could thus score between 0 and 20 on pragmatic (in)competence.

### Behavioral results

Fig. 6 shows the proportion of Expected (vs. Unexpected) referent choices and completion latencies in each condition for Experiment 2.

Comparing the left panel of Fig. 6 to the left panel in Fig. 2, the distribution of Expected and Unexpected discourse completions over conditions seems highly similar between experiments. A logistic mixed-effects analysis shows that all conditions again significantly differ from the *High-Adverb* condition: the probability of Expected referent clicks is significantly higher in the *High-Inderdaad* condition ( $\beta = 2.14$ ,

$SE = 0.51$ ,  $p < 0.001$ ), and significantly lower in the *High-Eigenlijk* condition ( $\beta = -3.03$ ,  $SE = 0.26$ ,  $p < 0.001$ ) as well as the *Medium-Eigenlijk* condition ( $\beta = -2.94$ ,  $SE = 0.26$ ,  $p < 0.001$ ).

In order to assess whether the presence of *eigenlijk* in medium-constraining contexts affected participants' preferred discourse completions, we compared click behavior in medium-constraining contexts between experiments (*Medium-Adverb* in Experiment 1 vs. *Medium-Eigenlijk* in Experiment 2). Results showed no significant effect of *Experiment* ( $\beta = -0.31$ ,  $SE = 0.19$ ,  $p = 0.11$ ), indicating that completions of medium-constraining dialogues did not change by the presence of *eigenlijk*. When comparing click behavior between experiments in the *High-Eigenlijk* condition, analysis revealed no significant effect of *Experiment* either ( $\beta = 0.49$ ,  $SE = 0.38$ ,  $p = 0.20$ ), suggesting that the increased frequency of non-referential *eigenlijk* in the experimental environment did not modulate the proportion of referential vs. non-referential interpretations of *eigenlijk* in high-constraining contexts.

The response latencies (right panel) show a pattern very similar to Experiment 1 as well. Unexpected discourse completions again seem slower than expected discourse completions, with the exception of medium-constraining contexts (where both referents are equally likely discourse completions). Linear mixed-effects analyses indeed showed a significant main effect of *Selected Referent* ( $\beta = -0.21$ ,  $SE = 0.11$ ,  $p < 0.05$ ) and a marginally significant *Selected Referent* by *Condition* interaction effect (comparing models with vs. without interaction effect:  $X^2(3) = 6.55$ ,  $p = 0.09$ ). Again, follow-up analyses revealed no significant differences between conditions for Unexpected discourse completions (all  $p$ 's  $> 0.48$ ); for Expected discourse completions, responses in the *High-Eigenlijk* condition were longer than in the *High-Adverb* condition ( $\beta = 0.29$ ,  $SE = 0.09$ ,  $p < 0.01$ ). Latencies did not differ between the *High-Inderdaad* condition and the *High-Adverb* condition ( $\beta = -0.13$ ,  $SE = 0.07$ ,  $p = 0.10$ ), but expected discourse completions were faster after encountering *inderdaad* than after encountering *eigenlijk* ( $\beta = -0.41$ ,  $SE = 0.09$ ,  $p < 0.001$ ).

Recall that in Experiment 1, Expected discourse completions were faster after *inderdaad* than after an adverb; however, this difference did not reach significance in Experiment 2. Conversely, Expected discourse completions after *eigenlijk* were slower than after an adverb in Experiment 2, whereas this difference did not reach significance in Experiment 1. A combined analysis of Expected discourse completion latencies across Experiments revealed that both DPs significantly differed from the *High-Adverb* condition (*High-Inderdaad*:  $\beta = -0.20$ ,  $SE = 0.06$ ,  $p < 0.01$ ; *High-Eigenlijk*:  $\beta = 0.21$ ,  $SE = 0.07$ ,  $p < 0.01$ ;

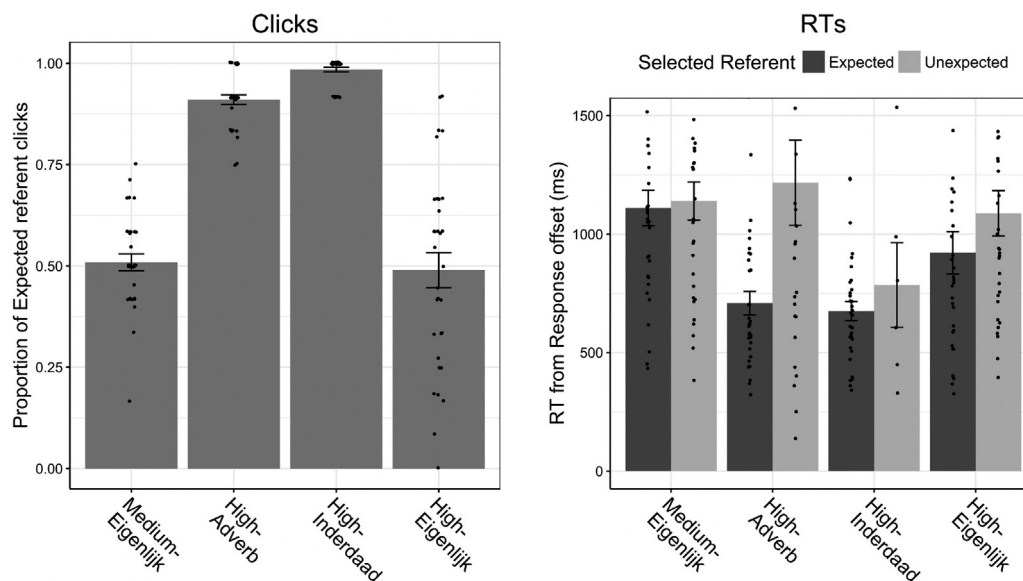


Fig. 6. Proportion of Expected (vs. Unexpected) referent choices (left panel) and response latencies (right panel) per condition in Experiment 2.

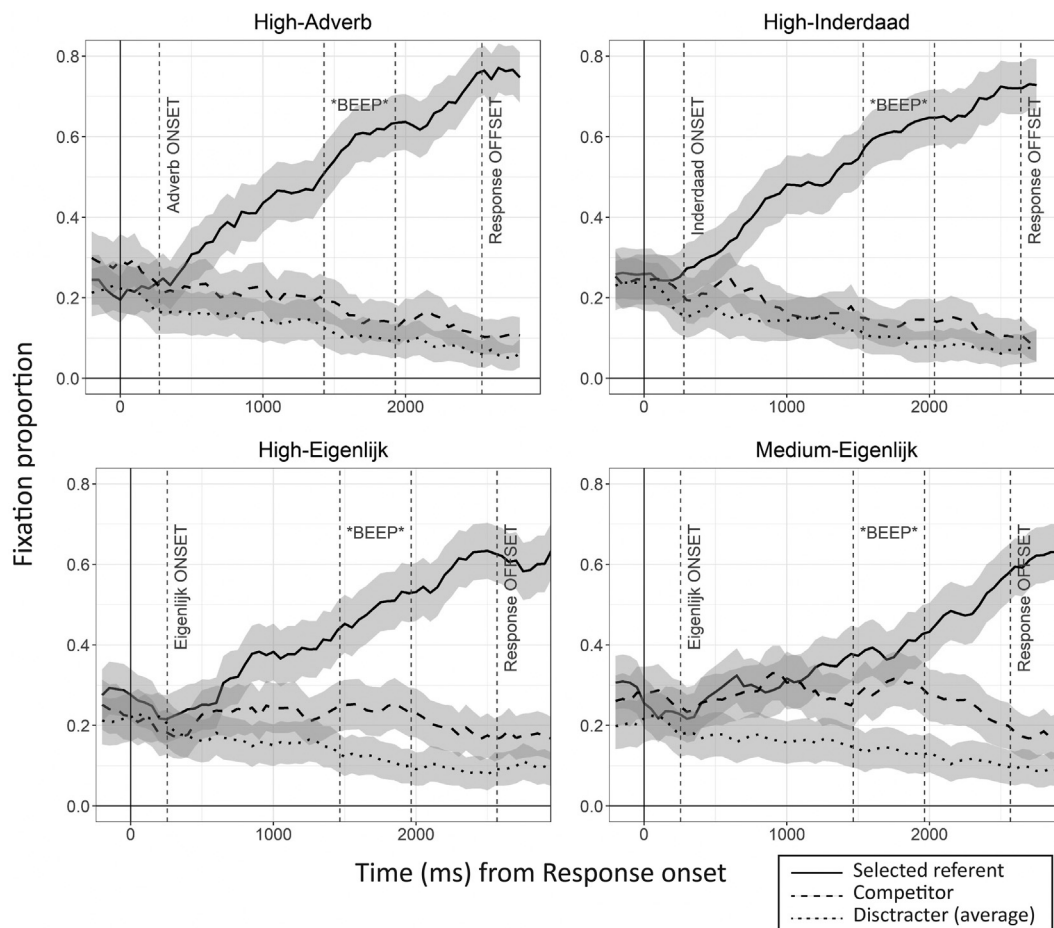


Fig. 7. Click-contingent gaze patterns over time (Response onset – Response offset) in each condition (Experiment 2).

*Medium-Adverb*:  $\beta = 0.46$ ,  $SE = 0.06$ ,  $p < 0.001$ ). We found no evidence for a main effect of *Experiment* ( $\beta = 0.15$ ,  $SE = 0.10$ ,  $p = 0.15$ ), nor for an *Experiment* by *Condition* interaction (comparing models with vs. without interaction effect:  $X^2(3) = 5.81$ ,  $p = 0.12$ ), allowing us to conclude that both expectation-managing DPs affect Expected dialogue completion latencies in opposite directions when compared with adverbs.

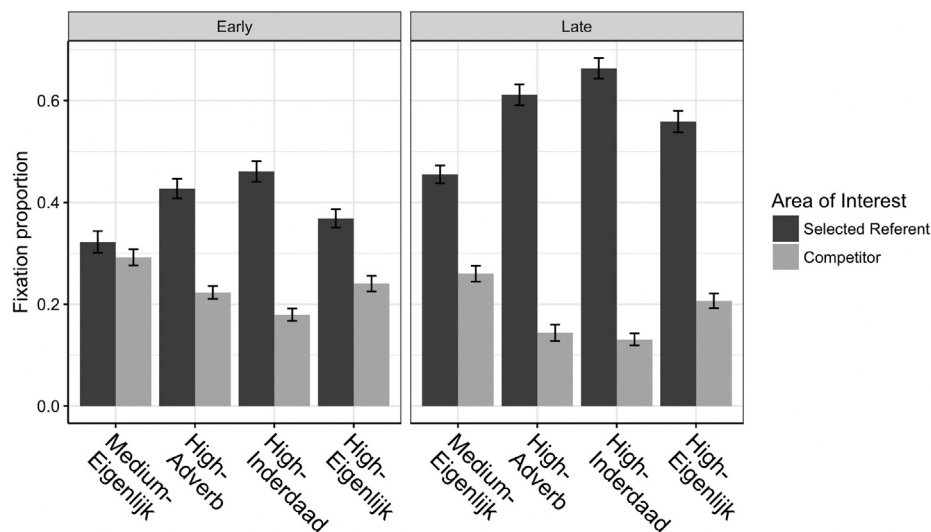
A comparison of reaction times in the *High-Eigenlijk* condition between experiments revealed a significant main effect of *Selected Referent* ( $\beta = -0.18$ ,  $SE = 0.08$ ,  $p < 0.05$ ), indicating that after encountering *eigenlijk*, Expected discourse completions were faster than Unexpected discourse completions. We found no evidence for a main effect of *Experiment* ( $\beta = 0.03$ ,  $SE = 0.12$ ,  $p = 0.74$ ), nor for an interaction effect ( $\beta = -0.16$ ,  $SE = 0.13$ ,  $p = 0.22$ ), suggesting that the increased frequency of *eigenlijk* in the experimental context neither speeded up nor slowed down participants' dialogue completions. Comparing response latencies in the *Medium-Eigenlijk* condition (Experiment 2) with the *Medium-Adverb* condition (Experiment 1), analyses showed no significant main effects of *Selected Referent* ( $\beta = -0.01$ ,  $SE = 0.06$ ,  $p = 0.85$ ) or *Experiment* ( $\beta = -0.05$ ,  $SE = 0.10$ ,  $p = 0.64$ ), nor a significant interaction effect ( $\beta = -0.02$ ,  $SE = 0.11$ ,  $p = 0.88$ ); this suggests that encountering *eigenlijk* instead of an adverb did not affect completion latencies in medium-constraining conversations.

As in Experiment 1, we found considerable variation across participants in the *High-Eigenlijk* condition. In order to assess whether these individual interpretation biases of *eigenlijk* were related to pragmatic competence, we included the individual AQ scores as predictor in the regression models. Participants' AQ scores were on average very low ( $M = 1.83$ ), indicating high pragmatic competence; the variation

between participants was relatively small ( $SD = 1.85$ ), with individual scores ranging from 0 to 8. Given the limited variation, it is not surprising that pragmatic competence did not significantly modulate the probability of Expected over Unexpected clicks in this condition ( $\beta = -0.07$ ,  $SE = 0.14$ ,  $p = 0.62$ ). Similarly, response latencies were not significantly modulated by pragmatic competence ( $\beta = -0.01$ ,  $SE = 0.04$ ,  $p = 0.86$ ), neither for Expected nor for Unexpected discourse continuations (comparing models with vs. without *Pragmatic competence* by *Selected Referent* interaction:  $X^2(2) = 0.14$ ,  $p = 0.71$ ). The lack of variation in pragmatic competence in our sample keeps us from drawing any conclusions about these null findings.

In sum, the behavioral results from Experiments 1 and 2 together confirm our hypothesis that expectation-managing DPs modulate comprehenders' expectations about the continuation of a dialogue. Encountering *inderdaad* led to more and faster Expected dialogue completions relative to an adverb; encountering *eigenlijk* yielded variation in dialogue completions. In addition, encountering *eigenlijk* slowed down dialogue completions when compared with an adverb. Unexpected completions were thereby more delayed than Expected completions, providing additional support for the assumption that integrating information in the visual context with a revised discourse model incurs additional processing costs.

We found no evidence that the increased frequency of non-referential *eigenlijk* modulated its preferred interpretation in high-constraining contexts, nor that encountering *eigenlijk* instead of an adverb affected the interpretation of medium-constraining dialogues. Because of the lack of individual variation in pragmatic competence, the question whether individual interpretation biases of interpersonal DPs are related to pragmatic abilities could not be answered successfully.



**Fig. 8.** Proportions of looks to the selected referent (dark grey bars) and the competitor referent (light grey bars) per condition during the Early (left panel) and Late Integration window (right panel) (Experiment 2).

#### Click-contingent gaze patterns

We performed the same analyses as in Experiment 1, comparing click-contingent AoI fixations between conditions in the Early and Late Integration time windows. In addition, we compared gaze patterns between Experiments.

We expected gaze patterns to be similar in the *High-Adverb* and *High-Inderdaad* condition, as these were the same across experiments. We were especially interested in differences between experiments in the *Medium* condition (to what extent does the presence of *eigenlijk* modulate gaze patterns in less constraining contexts?) and in the *High-Eigenlijk* condition (to what extent does the increased frequency of *eigenlijk* in the experiment modulate its effect on gaze behavior in strongly constraining contexts?).

Fig. 7 shows the proportion of click-contingent AoI fixations over time (from Response onset until Response offset) per condition; Fig. 8 represents the fixation proportions of selected referents and competitors per condition in the Early Integration (left panel) and Late Integration window (right panel).

Comparing the panels in Fig. 7 to those in Fig. 3, the gaze patterns look quite similar across conditions, with the exception of the medium-constraining conversations. Fixation proportions of Selected referents and Competitors seem to start diverging later in Experiment 2, suggesting that the presence of *eigenlijk* in the Response delayed participants' attentional preference for a specific interpretation in medium-constraining contexts.

Results of a linear mixed effects analysis showed that the preference to fixate the Selected referent over the Competitor in the *Medium-Eigenlijk* condition was not significant in the Early integration window ( $\beta = 0.42$ ,  $SE = 0.43$ ,  $p = 0.33$ ). In Experiment 1 we did find a significant preference to fixate the to-be-selected referent in the *Medium-Adverb* condition during the Early Integration window. This suggests that encountering *eigenlijk* in medium-constraining contexts slows down attentional preferences. However, a combined analysis of medium-constraining contexts alone showed no significant difference between experiments in the Early Integration window ( $\beta = 0.72$ ,  $SE = 0.49$ ,  $p = 0.15$ ). Hence, we have no evidence that attentional preferences to potential discourse referents in medium-constraining contexts were delayed by the presence of *eigenlijk* instead of an adverb. In the Late integration window, the attentional preference for the Selected referent over the Competitor is significant in the *Medium-Eigenlijk* condition ( $\beta = 2.76$ ,  $SE = 0.45$ ,  $p < 0.001$ ); again, an omnibus analysis of medium-constraining contexts revealed no significant difference

between experiments ( $\beta = -0.13$ ,  $SE = 0.57$ ,  $p = 0.82$ ).<sup>3</sup>

Next, we analyzed effects of expectation-managing DPs on visual attention in high-constraining conversations. The preference to fixate the selected referent was significantly stronger after encountering *inderdaad* than after encountering *eigenlijk* in both time windows (*Early*:  $\beta = 1.94$ ,  $SE = 0.50$ ,  $p < 0.001$ ; *Late*:  $\beta = 2.66$ ,  $SE = 0.57$ ,  $p < 0.001$ ). The stronger preference to fixate selected referents in the *Inderdaad*-condition compared to the *Adverb*-condition was marginally significant in both time windows (*Early*:  $\beta = 0.93$ ,  $SE = 0.47$ ,  $p = 0.06$ ; *Late*:  $\beta = 1.05$ ,  $SE = 0.60$ ,  $p = 0.09$ ). The preference to fixate the eventually selected referent after encountering *eigenlijk* was significantly weaker than after encountering an adverb (*Early*:  $\beta = -1.01$ ,  $SE = 0.48$ ,  $p < 0.05$ ; *Late*:  $\beta = -1.62$ ,  $SE = 0.58$ ,  $p < 0.01$ ).

In Experiment 1, we found no evidence for a difference between either DP condition and the *High-Adverb* condition in the Early Integration window; in Experiment 2, we did find a (tendency for a) difference. Analyses of the combined data revealed an immediate effect of both *inderdaad* and *eigenlijk* on visual attention when compared with the *Adverb*-condition (*Inderdaad*:  $\beta = 0.84$ ,  $SE = 0.30$ ,  $p < 0.01$ ; *Eigenlijk*:  $\beta = -0.85$ ,  $SE = 0.33$ ,  $p < 0.05$ ); both effects increased in the Late Integration time window (*Inderdaad*:  $\beta = 1.04$ ,  $SE = 0.36$ ,  $p < 0.01$ ; *Eigenlijk*:  $\beta = -1.81$ ,  $SE = 0.40$ ,  $p < 0.001$ ). Thus, findings from both experiments together confirm that listeners immediately integrate expectation-managing DPs to guide their visual attention during listening.

To assess whether the increased frequency of *eigenlijk* in Experiment 2 affects visual attention in the *High-Eigenlijk* condition, we compared gaze patterns in this condition between experiments. Results showed no significant effect of *Experiment* in either time window (*Early*:  $\beta = 0.11$ ,  $SE = 0.52$ ,  $p = 0.83$ ; *Late*:  $\beta = -1.00$ ,  $SE = 0.62$ ,  $p = 0.11$ ), suggesting that the presence of *eigenlijk* in medium-constraining contexts did not modulate its effect on attentional preferences in high-constraining contexts.

As in Experiment 1, we also assessed to what extent gaze patterns in the *High-Eigenlijk* condition are dependent on the final interpretation. Fig. 9 shows the click-contingent gaze patterns separately for Expected (left) and Unexpected (right) referent clicks in the *High-Eigenlijk* condition. The lines representing the Selected Referent (solid) and the Competitor (dashed) seem to come closer together in the second half of the response when the Unexpected referent is selected (right panel). This would suggest that semantic competition from contextually

<sup>3</sup> The final model included random intercepts for participants and items.

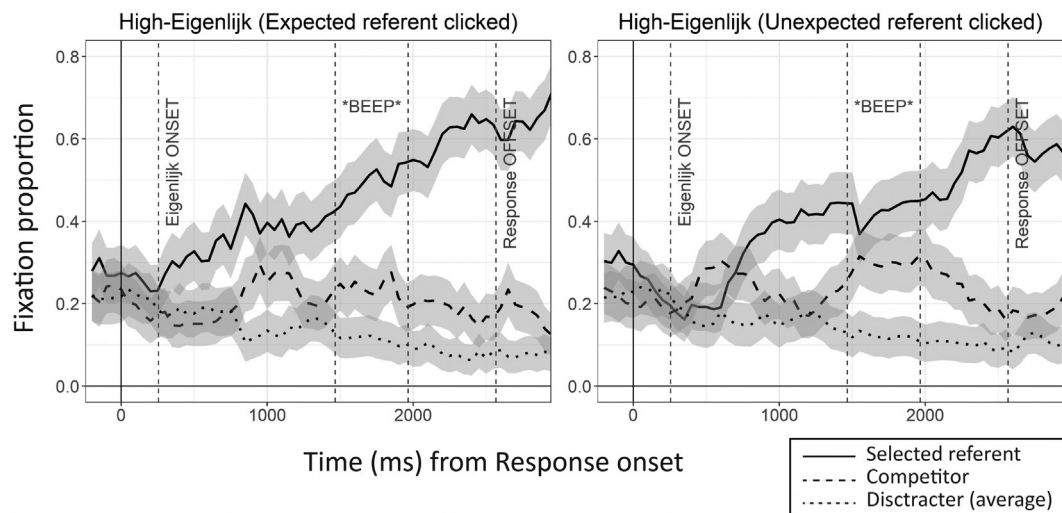


Fig. 9. Click-contingent gaze patterns over time for Expected (left) and Unexpected (right) referent clicks in the High-Eigenlijk condition (Experiment 2).

expected referents is stronger than from contextually unexpected referents.

Results from a linear mixed-effects regression analysis confirm an early preference to fixate the to-be-clicked referent ( $\beta = 1.30$ ,  $SE = 0.38$ ,  $p < 0.001$ ), but there was no difference between Expected and Unexpected referents in either of the time windows (*Early*:  $\beta = 1.07$ ,  $SE = 0.92$ ,  $p = 0.26$ ; *Late*:  $\beta = 1.13$ ,  $SE = 0.89$ ,  $p = 0.19$ ). In addition, omnibus analyses revealed that the interaction with *Experiment* is not significant in either time window (model comparison, *Early*:  $X^2(1) = 0.41$ ,  $p = 0.52$ ; *Late*:  $X^2(1) = 1.64$ ,  $p = 0.20$ ). Thus, results provide no evidence that the effect of *eigenlijk* on attentional preferences in high-constraining dialogues is modulated by the increased frequency of non-referential *eigenlijk* in the experiment.

This was corroborated by analyses of high-constraining trials in which the Expected referent was clicked. Results show that the preference to fixate Expected referents is stronger after encountering *inderdaad* than after encountering *eigenlijk* in both time windows (*Early*:  $\beta = 1.44$ ,  $SE = 0.59$ ,  $p < 0.05$ ; *Late*:  $\beta = 2.32$ ,  $SE = 0.82$ ,  $p < 0.01$ ). There was no evidence for a difference between *inderdaad* and adverbs in Expected referent fixations (*Early*:  $\beta = 0.79$ ,  $SE = 0.48$ ,  $p = 0.11$ ; *Late*:  $\beta = 0.71$ ,  $SE = 0.63$ ,  $p = 0.27$ ). There was a tendency to fixate Expected referents less after encountering *eigenlijk* relative to an adverb in the Late Integration window,  $\beta = -1.61$ ,  $SE = 0.82$ ,  $p = 0.06$ , but not in the Early Integration window,  $\beta = -0.63$ ,  $SE = 0.63$ ,  $p = 0.33$ . Analysis of the two experiments together reveal a stronger preference to fixate Expected referents after hearing *inderdaad* than after encountering an adverb in both time windows (*Early*:  $\beta = 0.73$ ,  $SE = 0.33$ ,  $p < 0.05$ ; *Late*:  $\beta = 0.89$ ,  $SE = 0.34$ ,  $p < 0.05$ ).<sup>4</sup> Expected referents tended to be fixated less after encountering *eigenlijk* relative to an adverb in the Early integration window ( $\beta = -0.67$ ,  $SE = 0.40$ ,  $p = 0.09$ ); the difference was stronger in the Late Integration window ( $\beta = -2.01$ ,  $SE = 0.56$ ,  $p < 0.001$ ). We found no evidence for a main effect of *Experiment*, nor for an *Experiment* by *Condition* interaction effect in either time window (all  $p$ 's  $> 0.25$ ). Together, these results provide suggestive evidence that both expectation-managing DPs immediately affected the preference to fixate Expected referents if these are eventually selected as dialogue continuations, but no evidence that effects of *eigenlijk* on attentional preferences dialogues are modulated by the increased frequency of non-referential *eigenlijk* in Experiment 2.

We also compared Unexpected click-contingent fixations in the *High-Eigenlijk* condition with Expected click-contingent fixations in the

other two conditions. For both time windows, results showed a weaker preference to fixate Unexpected referents after *eigenlijk* when compared with the preference to fixate Expected referents after *inderdaad* (*Early*:  $\beta = -2.61$ ,  $SE = 0.75$ ,  $p < 0.01$ ; *Late*:  $\beta = 3.05$ ,  $SE = 0.85$ ,  $p < 0.001$ ), as well as after adverbs (*Early*:  $\beta = -1.82$ ,  $SE = 0.79$ ,  $p < 0.05$ ; *Late*:  $\beta = -1.99$ ,  $SE = 0.72$ ,  $p < 0.01$ ). A combined analysis of both experiments confirms that across experiments, Unexpected referents were fixated less after encountering *eigenlijk* when compared with Expected referent fixations after encountering *inderdaad* (*Early*:  $\beta = -2.13$ ,  $SE = 0.43$ ,  $p < 0.001$ ; *Late*:  $\beta = -3.27$ ,  $SE = 0.51$ ,  $p < 0.001$ ) or an adverb (*Early*:  $\beta = -1.42$ ,  $SE = 0.45$ ,  $p < 0.01$ ; *Late*:  $\beta = -2.32$ ,  $SE = 0.57$ ,  $p < 0.001$ ). Thus, if participants commit to a referentially contrastive interpretation of *eigenlijk*, our findings show that *eigenlijk* immediately increases competition from alternative dialogue interpretations. Again, there were no main effects of *Experiment*, nor *Experiment* by *Condition* interaction effects in either time window (all  $p$ 's  $> 0.24$ ).

In sum, findings from both experiments together confirm that both *inderdaad* and *eigenlijk* immediately affect visual attention to possible dialogue continuations, when compared with an adverb, in opposite directions. We found no evidence that the increased frequency of non-referential *eigenlijk* in Experiment 2 influenced effects of *eigenlijk* on visual attention in high-constraining contexts, nor that the presence of *eigenlijk* modulated visual attention to discourse referents in medium-constraining dialogues.

#### Exploratory analysis: voice-based adaption

We found no evidence that an increase in non-referential uses of *eigenlijk* (in Experiment 2) modulated interpretations of *eigenlijk* in high-constraining contexts, despite ample evidence in the literature showing that comprehenders' previous language experience influences their processing of subsequent linguistic input. However, our experimental manipulation may have been too general to capture adaptation effects. That is, prior research has shown that comprehenders keep track of talker-specific idiosyncrasies in various language domains (e.g., phonological, syntactic, pragmatic), and use this knowledge during language processing (e.g., Arnold et al., 2007, van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008; Bosker et al., 2014; Eisner & McQueen, 2005; Grodner & Sedivy, 2011; Kamide, 2012; Kraljic & Samuel, 2005; Trude & Brown-Schmidt, 2011). Recall that the dialogues in our experiments were recorded by two male and two female voices. Maybe participants, rather than tracking uses of *eigenlijk* across the board, may have used speaker voice information to track speaker-specific uses of *eigenlijk*.

In Experiment 1, male uses of *eigenlijk* (64%) were more often

<sup>4</sup> Final models included random intercepts for participants and items, as well as random by-item slopes for *Condition*.

interpreted referentially than female uses of *eigenlijk* (52%); this difference was much smaller in Experiment 2 (53% vs. 50%, respectively). Note that speaker voice was not counterbalanced within each experiment, male and female voices were associated with different stimuli. Across experiments, however, stimuli in the *High-Eigenlijk* condition were identical, and thus recorded by the same voices. This allowed for an exploratory analysis of dialogue completions in the *High-Eigenlijk* condition, including Voice (male vs. female), Experiment, and their interaction as independent predictors. Results revealed an Experiment by Voice interaction (comparing models with vs. without the interaction effect:  $X^2(2) = 4.04, p < 0.05$ ). Follow-up analyses revealed that, relative to Experiment 1, participants were less inclined to interpret male uses of *eigenlijk* referentially (i.e., fewer Unexpected dialogue completions) in Experiment 2 (where the frequency of non-referential uses of *eigenlijk* increased) (difference 11%;  $\beta = -0.87, SE = 0.45, p = 0.05$ ). By contrast, no difference between Experiments was found for female uses of *eigenlijk* (difference 2%;  $\beta = -0.09, SE = 0.40, p = 0.82$ ).

These results provide suggestive evidence that comprehenders were sensitive to talker-specific idiosyncrasies in *eigenlijk* use, adapting their behavior differentially for male and female talkers. Since we have too few data points for further explorations of voice effects separately for Unexpected and Expected clicks, we will leave this for future research.

## General discussion

The present study set out to investigate how language comprehenders use interpersonal discourse particles in online discourse understanding. The combination of offline and online measures enabled us to examine distinct stages of the interpretation process: we identified how and when effects of interpersonal discourse particles on interpretation arise, thereby further specifying and extending theoretical claims about their conversation-managing function. Moreover, by allowing for variability in interpretations in our experimental set-up, we embraced and exploited the notorious polyfunctionality of discourse markers, showing that listeners make rapid and fine-grained inferences about the use of interpersonal discourse particles in conversational interaction.

Encountering *inderdaad* quickly reduced listeners' attention to alternative discourse continuations during listening, and yielded more and faster contextually likely dialogue completions relative to encountering an adverb. A linguistic confirmation of inferred expectedness thus proves beneficial for the listener: it strengthens expectations about the unfolding input, which reduces the amount of cognitive resources spent on distributing probabilities over alternative interpretations, which in turn speeds up the interpretation process. Conversely, encountering *eigenlijk* decreased the likelihood of a contextually expected discourse continuation, as reflected in increased visual attention to alternative interpretations during listening, and more contextually contrastive dialogue completions. If *eigenlijk* was interpreted referentially, we found that listeners quickly revised their initial expectation, as reflected in an immediate attention shift from the contextually expected to the contextually contrastive referent. This finding shows that the inferred content was readily available to the listener, supporting the view that pragmatic inferences can be generated immediately (e.g., Grodner & Sedivy, 2010; Levinson, 2000). However, this early commitment to a referential interpretation of *eigenlijk* did not result in faster dialogue completions: participants experienced sustained visual competition from alternative dialogue interpretations, and were slower to complete dialogues with the referent that contrasted with the contextually most likely interpretation. This finding is consistent with studies reporting lingering predictions or interpretations, as measured by sustained visual attention to initial interpretations that need to be revised as the linguistic input unfolds (e.g., Corley, 2010; Lowder & Ferreira, 2016). Moreover, it provides further support for the view that delays in processing of implicatures reflect additional processing costs associated with integrating the interpretation of a pragmatic inference

with relevant information in the context, rather than the costs associated with computing a pragmatic inference itself (Grodner et al., 2010; Kurumada et al., 2014).

Behavioral results showed that *eigenlijk* was not always interpreted in the same way: only about half of the time, *eigenlijk* was interpreted as encoding a contrast with a referential expectation. In the other half of the cases, listeners completed dialogues with the contextually most likely referent. Findings from gaze patterns and completion latencies reject the possibility that *eigenlijk* was ignored in these cases: relative to encountering an adverb, encountering the contrastive cue still increased visual competition from alternative dialogue interpretations, and yielded delayed completions. As such, our findings underline the notorious polyfunctionality of discourse markers. With a more traditional paradigm (i.e., with the referent mentioned in the Response rather than replaced by a beep), non-referential interpretations of *eigenlijk* would have been experimentally excluded. This may induce an experimental interpretation bias in trials with *eigenlijk*, thereby decreasing the ecological validity of our study. Moreover, if dialogues with an adverb or *inderdaad* unambiguously ended with the Expected referent, we would have experimentally increased the probability of a contextually likely discourse continuation in these conditions. As a result, the additional confirmation from *inderdaad* may have no longer been beneficial to the listener, and the difference between *inderdaad* and adverbs may have disappeared. Notwithstanding the informativity of such findings with respect to adaptation during language processing, we believe that our current findings, obtained with the adapted paradigm, provide a more realistic picture of the function of DPs in everyday conversational interaction, including listeners' uncertainty about how a dialogue will continue, as well as speakers' use of both referential and non-referential *eigenlijk*.

The between-participant variation in behavioral responses suggested that variable interpretations of *eigenlijk* are not only discourse-dependent, but also listener-specific. This finding has important implications for linguistic research on DPs. Both qualitative and quantitative analyses of DPs involve in-depth investigations of the conversational circumstances under which DPs were used by the speaker, from which researchers infer their communicative intentions. Even in approaches that explicitly acknowledge the interactional aspect of DP use, researchers infer the ways in which DPs are understood by the addressee from the conversational context in which they were produced (for instance, by examining the continuation of the conversation). Whereas such analyses are highly valuable for a better understanding of DPs, our empirical findings show that DPs can be interpreted differently by different listeners even if they occur in exactly the same conversational context. As such, we believe that experimental investigations of their role in discourse understanding are necessary for a full account of the function of DPs in language.

In Experiment 2, we further explored the nature of individual variation in interpretations of *eigenlijk*. We concentrated on two factors potentially involved in individual interpretation biases, namely (1) the frequency distribution of referential and non-referential uses of *eigenlijk* encountered previously, and (2) individual differences in pragmatic competence. Unfortunately, with respect to this last factor, we were unable to test theory-driven hypotheses because of the lack of variation in AQ scores in our sample. More promising results are to be expected from studies comparing DP interpretation in more heterogeneous populations in terms of pragmatic competence, for instance, in individuals that are clinically diagnosed with autism spectrum disorder. This being said, the fact that we found considerable inter-individual variation in interpretations of *eigenlijk*, despite relative invariance in inter-individual pragmatic competence, suggests that pragmatic competence cannot be the sole determinant in explaining variation in interpretations of *eigenlijk*.

With respect to the first factor, initial analyses provided no evidence that the increased frequency of non-referential uses of *eigenlijk* in the experiment affected interpretations of *eigenlijk* in high-constraining

contexts. However, results from an exploratory analysis suggest that participants were in fact sensitive to frequency distributions of (non-) referential *eigenlijk* use at a more fine-grained level. In Experiment 1, male uses of *eigenlijk* were more often interpreted referentially than female uses of *eigenlijk*, possibly suggesting that male speakers more often use referential *eigenlijk* than female speakers in everyday conversational interaction. Results from Experiment 2 provide suggestive evidence for voice-based adaptation in pragmatic language processing: participants were less likely to interpret male uses of *eigenlijk* referentially if their frequency of non-referential use of *eigenlijk* increased, whereas no such adaptation effect was found for female voices. We speculatively relate this voice-based adaptation to a stereotypical difference between men and women, namely the ability to empathize with another individual. Previous research has shown that women generally score higher than men on various self-assessed empathy scales (e.g., Eisenberg & Lennon, 1983); moreover, empathy has been related to social language abilities (e.g., Baron, 1995; van den Brink et al., 2012). On the basis of this stereotypical information, participants in Experiment 1 may have inferred that male speakers are less sensitive to the social characteristics of the dialogue, and hence interpreted their use of *eigenlijk* more referentially. In Experiment 2 (where the proportion of non-referential uses was increased for all talkers), participants may have adjusted these stereotypical ideas about male usage of *eigenlijk*, resulting in less referential interpretations of male uses of *eigenlijk*. Although this analysis was not planned *a priori* and the results are in need of replication, findings suggest that situation-specific interpretations of *eigenlijk* also include inferences about speaker-specific knowledge. Future research could investigate to what extent speaker-specific variation in DPs use affects online processing of DPs, which the current data was not designed for. It could for instance reveal whether comprehenders more quickly commit to a referential interpretation of *eigenlijk* if produced by a male rather than a female speaker.

Given the intersubjective meanings of *inderdaad* and *eigenlijk*, our findings contribute to the evidence that common ground and perspective-taking have immediate effects on incremental language processing, in line with previous findings from visual world experiments involving interactive tasks (e.g., Brown-Schmidt & Tanenhaus, 2008; Hanna et al., 2003). The current experiment did not involve an interactive paradigm, but we showed that listeners are immediately sensitive to the interpersonal meanings of *eigenlijk* and *inderdaad* even when performing a task as overhearers. This is consistent with findings from Tolins and Fox Tree (2016), who showed that addressee backchannels influenced overhearers' processing of unfolding dialogues; in addition, our results provide further evidence that overhearers can generate predictions about subsequent input over the turns of multiple speakers (see also Bögels, Kendrick, & Levinson, 2015). At the same time, "effects of perspective are likely to be strongest in tasks where participants have joint goals, common ground is established collaboratively, and exchange of information is negotiated by both parties" (Brown-Schmidt et al., 2008, p. 1133). As such, it remains an open question whether comprehenders would be differentially sensitive to interpersonal DP uses if they actively engage in a conversation. In an fMRI study comparing face-saving indirect speech act processing in addressees and overhearers, Bašnáková, van Berkum, Weber, & Hagoort (2015) found similar activation of neural networks associated with cognitive perspective-taking in addressees and overhearers. Based on this finding, the authors argue that "cognitive perspective-taking presumably occurs with any contextualized utterance, for the simple reason that listeners will want to know what the speaker is *really* talking about, and what his or her social intentions are" (p. 87). In addition, they found increased activation of networks associated with emotional salience and empathy in addressees relative to overhearers, suggesting increased emotional involvement in face-saving indirect speech processing if comprehenders are directly addressed. Future research could investigate to what extent increased emotional involvement may influence interpretation of face-saving uses of interpersonal DPs in a more interactive task.

Empirically, because of their immediate effects on discourse processing, we believe that interpersonal DPs provide useful tools to further investigate the nature of discourse-based predictions in incremental language processing. For instance, expectation-managing DPs can be used to investigate to what extent processing of (un)expected input is modulated if a contextual prediction is made more or less likely. In the current paradigm, the disambiguating information was never encountered: participants completed dialogues with the referent they believed was most likely. Hence, it remains an open question whether *eigenlijk* is useful to the comprehender to the extent that its presence facilitates processing of subsequent unexpected input relative to a control condition. Moreover, the visual world paradigm limits possible interpretations to the options presented in the visual display, whereas comprehenders' discourse models may contain many more potential interpretations (for discussion, see e.g., Huettig, Rommers, & Meyer, 2011). Future research could use EEG methodology to investigate whether and to what extent expectation-managing DPs facilitate incremental processing of subsequent (un)expected input.

The present study specifically focused on two discourse particles in Dutch, but many languages exhibit discourse markers that encode largely comparable basic meanings. For instance, the basic contrastive/adversative meaning of *eigenlijk* has also been identified for German *eigentlich* (Eckardt, 2009), English *actually/in fact* (e.g., Aijmer, 2013), French *en fait* (e.g., Mortier & Degand, 2009) and Chinese *qishi/shishi-shang* (Wang, Tsai, & Yang, 2010). At the same time, however, discourse markers typically lack clear translation equivalents, implying that the range of possible functions associated with an individual discourse marker varies greatly across languages. For instance, a parallel corpus study by Mortier and Degand (2009) showed that not all uses of Dutch *eigenlijk* are translated with their closest French equivalent *en fait*, nor vice versa. Strikingly, almost half of the uses of Dutch *eigenlijk* was left untranslated in French, suggesting that Dutch *eigenlijk* has a wider range of uses than its French counterpart(s), which may speculatively indicate a further stage of grammaticalization (e.g., Traugott, 2010; Traugott & Dasher, 2002). Our findings showed that listeners immediately integrated the contrastive information encoded in *eigenlijk* to reduce their visual attention to contextually likely discourse referents, irrespective of the attested variation in dialogue completions. Based on these findings, we would predict discourse markers with similar basic meanings to have similar effects on visual attention, regardless of their full functional range.

Finally, the adapted version of the visual world paradigm introduced here offers a unique opportunity to empirically assess cross-linguistic variation in interpretations of contrastive discourse markers. We may, for instance, expect a smaller proportion of non-referential interpretations of contrastive discourse markers with a smaller functional range relative to Dutch *eigenlijk*. In addition, the adapted paradigm could be used to investigate variation in language learners' interpretations of *eigenlijk*. Given that interpersonal functions of discourse markers are notoriously difficult to learn for both first and second language learners (Fox Tree, 2010), we predict face-saving uses of *eigenlijk* to be hardest to acquire for learners; we may hence expect the proportion of non-referential interpretations of *eigenlijk* to be smaller in learners, but to increase with proficiency. Future comprehension studies comparing discourse markers across languages and in different populations may thus build on the present findings to advance our understanding of their polyfunctionality.

#### Acknowledgments

This research was supported by the Netherlands Organisation for Scientific Research (NWO, grant 275-89-022, awarded to the first author). We thank the editor-in-chief and three anonymous reviewers for valuable comments on earlier versions of this paper, as well as Vera van 't Hoff, Marlou Rasenberg and Anne Mickan for help with stimulus design and data acquisition.



## A. List of 10 stimuli with approximate translations

1	High	Medium
Context	Monique is zwanger, en komt haar vriendin Ine tegen in de kroeg. <i>Monique is pregnant, and runs into her friend Ine in a bar.</i>	Monique komt haar vriendin Ine tegen in een cafeetje in de stad. <i>Monique runs into her friend Ine in a little bar in the city.</i>
Question	Jij bent zeker de Bob vandaag? <i>You must be the designated driver today?</i>	Lust je ook iets te drinken? <i>Would you like a drink?</i>
Response	Ik stond net/inderdaad/eigenlijk op het punt een [cola/wijn] te bestellen. <i>I was just/indeed/actually about to order a [coke/wine].</i>	
2	High	Medium
Context	Puck en Pleun gaan bij Esmee op bezoek die met gym haar arm gebroken heeft. <i>Puck en Pleun are visiting Esmee who broke her arm during gymnastics class.</i>	Puck en Pleun gaan bij Esmee op bezoek die met gym haar arm gebroken heeft. <i>Puck en Pleun are visiting Esmee who broke her arm during gymnastics class.</i>
Question	Zullen we iets gezonds voor haar meenemen? <i>Shall we bring her something healthy?</i>	Zullen we namens de klas iets langsbrengen? <i>Shall we bring her something on behalf of the class?</i>
Response	Het lijkt me best/inderdaad/eigenlijk leuk om een [fruitmand/teddybeer] te geven. <i>I would quite/indeed/actually like to give her a [fruit basket/teddy bear].</i>	
3	High	Medium
Context	Martine en Nancy zitten een zielige film te kijken, als Martine ineens sniffend opstaat. <i>Martine and Nancy are watching a sad movie when Martine suddenly stands up sniffing</i>	Martine en Nancy zitten tv te kijken als Martine ineens opstaat. <i>Martine and Nancy are watching TV when Martine suddenly stands up.</i>
Question	Je wilt zeker je tranen wegvegen? <i>Do you want to go wipe your tears?</i>	Wat ga je doen? <i>What are you up to?</i>
Response	Ik ga even/inderdaad/eigenlijk een pakje [zakdoeken/sigaretten] halen. <i>I'm just/indeed/actually going to get a pack of [tissues/cigarettes].</i>	
4	High	Medium
Context	Het is winter en na schooltijd komt Tom Bart tegen in de fietsenstalling. <i>It is wintertime, and Tom runs into Bart at the bicycle stand after school.</i>	Bart en Tom hebben zich voorgenomen om meer te gaan bewegen. <i>Bart and Tom intend to exercise more.</i>
Question	Jij gaat toch ook mee naar de ijsbaan? <i>You are joining us to the ice rink, right?</i>	Ga je vandaag nog wat sportiefs doen? <i>Are you going to do something sporty today?</i>
Response	Ik heb best/inderdaad/eigenlijk zin om lekker te [schaatsen/fietsen] vanmiddag. <i>I quite/indeed/actually feel like [ice skating/cycling] this afternoon.</i>	
5	High	Medium
Context	Yvette komt lijkleek thuis van de kermis. <i>Yvette comes home from the fair looking white as a sheet.</i>	Yvette komt thuis en ziet lijkleek. <i>Yvette comes home looking white as a sheet.</i>
Question	Had je weer last van hoogtevrees? <i>Did your fear of heights act up again?</i>	Wat is er met jou aan de hand? <i>What is up with you?</i>
Response	Ik werd helemaal/inderdaad/eigenlijk niet goed in het [reuzenrad/spookhuis] vanmiddag. <i>I got really/indeed/actually unwell in the [Ferris wheel/haunted house] this afternoon.</i>	
6	High	Medium
Context	Na een concert in het stadion komt Loek Sabine tegen. <i>After a concert at the stadium, Loek runs into Sabine.</i>	Na een concert in het stadion komt Loek Sabine tegen. <i>After a concert at the stadium, Loek runs into Sabine.</i>
Question	Ga je ook met het openbaar vervoer terug naar de stad? <i>Are you also going back to the city by public transport?</i>	Hoe ga jij naar huis? <i>How are you going home?</i>
Response	Ik sta hier/inderdaad/eigenlijk al een tijdje op een ... te wachten. <i>I have here/indeed/actually been waiting for a [bus/taxi] for a little while.</i>	

	High	Medium
7		
Context	Op oudjaarsavond komt Fred met een auto vol dozen terug uit de stad. <i>On New Year's eve, Fred returns from the city with a car full of boxes.</i>	Fred en zijn buurman komen elkaar vlak na oud en nieuw tegen. <i>Fred and his neighbour run into each other right after New Year's day.</i>
Question	Zo te zien ga je weer knallend het jaar uit? <i>Looks like you are ending the year with a bang?</i>	Heb je ook een dure feestmaand gehad? <i>Did you have an expensive holiday month, too?</i>
Response	Ik heb dit jaar/inderdaad/eigenlijk voor een kapitaal aan [vuurwerk /oliebollen] uitgegeven. <i>I have this year/indeed/actually spent a fortune on [fireworks/Dutch New Year's snack].</i>	
8		
Context	Hanneke brengt een bezoek aan een brouwerij in de stad. <i>Hanneke is visiting a brewery in the city.</i>	Hanneke zit in een knus cafeetje in haar lievelingsstad. <i>Hanneke is in a cosy bar in her favourite city.</i>
Question	Lust je iets van de tap? <i>Would you like something on tap?</i>	Lust je nog iets te drinken? <i>Would you like another drink?</i>
Response	Ik wil ontzettend/inderdaad/eigenlijk graag jullie nieuwste [bier/cocktail] proberen. <i>I would really/indeed/actually like to try your newest [beer/cocktail].</i>	
9		
Context	Op een zonnige middag heeft Casper zijn moeder aan de telefoon. <i>On a sunny day Casper is on the phone with his mother.</i>	Op een middag heeft Casper zijn moeder aan de telefoon. <i>On an afternoon, Casper is on the phone with his mother.</i>
Question	Ga je straks buiten eten met dit mooie weer? <i>Are you going to eat outside with this beautiful weather?</i>	Heb je nog plannen vandaag? <i>Do you have any plans for today?</i>
Response	Ik ga gezellig/inderdaad/eigenlijk met vrienden [barbecueën/poolen] vanavond. <i>I will cozily/indeed/actually be [barbecuing/playing pool] with friends tonight.</i>	
10		
Context	De avond voor zijn zoontjes verjaardag komt Henk de trap afgestommeld. <i>The evening before his son's birthday, Henk comes stumbling down the stairs.</i>	De avond voor zijn zoontjes verjaardag komt Henk de trap afgestommeld. <i>The evening before his son's birthday, Henk comes stumbling down the stairs.</i>
Question	Je hebt zeker de versiering van zolder gehaald? <i>Were you getting the decorations from the attic?</i>	Wat was je op zolder aan het doen? <i>What were you doing in the attic?</i>
Response	Ik heb alvast/inderdaad/eigenlijk de doos met [slingers/kadootjes] gepakt. <i>I have already/indeed/actually taken the box with [garlands/presents].</i>	

## References

- Aijmer, K. (2002). *English discourse particles*. Amsterdam/Philadelphia: John Benjamins.
- Aijmer, K. (2013). *Understanding pragmatic markers: A variational pragmatic approach*. Edinburgh: Edinburgh University Press.
- Aijmer, K., & Simon-Vandenberg, A. M. (2004). A model and a methodology for the study of pragmatic markers: The semantic field of expectation. *Journal of Pragmatics*, 36, 1781–1805.
- Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–264.
- Arnold, J. E., Hudson Kam, C. L., & Tanenhaus, M. K. (2007). If you say *thee uh* you are describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 914–930.
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clucley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31, 5–17.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- Bašnáková, J., van Berkum, J., Weber, K., & Hagoort, P. (2015). A job interview in the MRI scanner: How does indirectness affect addressees and overhearers? *Neuropsychologia*, 76, 79–91.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models Using lme4. *Journal of Statistical Software*, 67, 1–48.
- van Bergen, G., van Gijn, R., Hogeweg, L., & Lestrade, S. (2011). Discourse marking and the subtle art of mind-reading: The case of Dutch *eigenlijk*. *Journal of Pragmatics*, 43, 3877–3892.
- van Bergen, G., & Flecken, M. (2017). Putting things in new places: Linguistic experience modulates the predictive power of placement verb semantics. *Journal of Memory and Language*, 92, 26–42.
- van Berkum, J. J., van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20, 580–591.
- Bögels, S., Kendrick, K. H., & Levinson, S. C. (2015). Never say no... How the brain interprets the pregnant pause in conversation. *PLoS One*, 10, e0145474.
- Bosker, H. R., Quené, H., Sanders, T. J. M., & de Jong, N. H. (2014). Native 'um' elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language*, 75, 104–116.
- van den Brink, D., van Berkum, J. J. A., Bastiaansen, M. C. M., Tesink, C. M. J. Y., Kos, M., Buitelaar, J. K., & Hagoort, P. (2012). Empathy matters: ERP evidence for inter-individual differences in social language processing. *Social, Cognitive and Affective Neuroscience*, 7, 173–182.
- Brodeur, M., Dionne-Dosti, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a New Set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE*, 5, e10773.
- Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: A targeted language game approach. *Cognitive Science*, 32, 643–684.
- Brown-Schmidt, S., & Hanna, J. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialogue & Discourse*, 2, 11–33.
- Brown-Schmidt, S., & Heller, D. (2018). Perspective-taking during conversation. In G. Gaskell, & S. A. Rueschmeyer (Eds.). *Oxford Handbook of Psycholinguistics* (2nd ed.). Oxford: Oxford University Press.
- Canestrelli, A., Mak, W. M., & Sanders, T. J. M. (2012). Causal connectives in discourse processing: How differences in subjectivity are reflected in eye movements. *Language and Cognitive Processes*, 28, 1394–1413.
- Clark, H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84, 73–111.
- Corley, M. (2010). Making predictions from speech with repairs: Evidence from eye movements. *Language and Cognitive Processes*, 25, 706–727.
- Corley, M., & Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of *um*. *Language and Linguistics Compass*, 2, 589–602.
- Corley, M., MacGregor, L. J., & Donaldson, D. I. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition*, 105, 658–668.
- Cozijn, R., Commandeur, E., Vonk, W., & Noordman, L. (2011). The time course of the use

- of implicit causality information in the processing of pronouns: A visual world paradigm study. *Journal of Memory and Language*, 64, 381–403.
- Degand, L., Cornillie, B., & Pietrandrea, P. (Eds.). (2013). *Discourse markers and modal particles: Categorization and description*. Amsterdam/Philadelphia: John Benjamins.
- Delaney-Busch, N., Morgan, E., Lau, E., & Kuperberg, G. (2018). Comprehenders rationally adapt semantic predictions to the statistics of the local environment: a Bayesian model of trial-by-trial N400 amplitudes. In E. Davelaar, G. Gunzelmann, H. Howes, & T. Tenbrink (Eds.). *Proceedings of the 39th Annual Conference of the Cognitive Science Society*. London: Cognitive Science Society.
- Diewald, G. (2010). On some problem areas in grammaticalization studies. In K. Stathi, E. Gehweiler, & E. König (Eds.). *Grammaticalization: Current views and issues (Studies in Language Companion Series 119)* (pp. 17–50). Amsterdam/Philadelphia: John Benjamins.
- Drenhaus, H., Demberg, V., Köhne, J., & Delogu, F. (2014). Incremental and predictive discourse processing based on causal and concessive discourse markers: ERP studies on German and English. In Proceedings of the 36th Annual Conference of the Cognitive Science Society (CogSci 2014).
- Eckardt, R. (2009). The real, the apparent, and what is eigentlich. *Oslo Studies in Language*, 1, 77–108.
- Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin*, 94, 100–131.
- Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Attention, Perception, & Psychophysics*, 67, 224–238.
- Fine, A., Jaeger, T. F., Farmer, T., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *Plos One*, 8, e77661.
- Finlayson, I. R., & Corley, M. (2012). Disfluency in dialogue: An intentional signal from the speaker? *Psychonomic Bulletin & Review*, 19, 921–928.
- Fischer, K. (2006). *Approaches to discourse particles (studies in pragmatics)*, Vol. 1. Amsterdam: Elsevier.
- Fischer, K. (2014). Discourse markers. In K. Schneider, & A. Barron (Vol. Eds.), *Pragmatics of discourse (handbooks of pragmatics): Vol. 3*, (pp. 271–294). Berlin: Mouton de Gruyter.
- Fox Tree, J. E., & Schrock, J. C. (1999). Discourse markers in spontaneous speech: Oh what a difference an Oh makes. *Journal of Memory and Language*, 40, 280–295.
- Fox Tree, J. E., & Schrock, J. C. (2002). Basic meanings of you know and I mean. *Journal of Pragmatics*, 34, 727–747.
- Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & Cognition*, 29, 320–326.
- Fox Tree, J. E. (2010). Discourse markers across speakers and settings. *Language & Linguistics Compass*, 4, 269–281.
- Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, 31, 931–952.
- Gerwien, J., & Rudka, M. (in press). Expectation changes over time: How long it takes to process focus imposed by German 'sogar'. In: I. Recio, L. Nadal, A. Cruz, & O. Loureda (Eds.), *Methodological approaches to discourse markers (Pragmatics and beyond series)*. Amsterdam/Philadelphia: John Benjamins.
- Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. L. Morgan (Eds.). *Syntax and Semantics, volume 3: Speech Acts* (pp. 41–58). New York: Academic Press.
- Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some", and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, 116, 42–55.
- Grodner, D., & Sedivy, J. (2011). The effect of speaker-specific information on pragmatic inferences. In E. A. Gibson, & N. J. Pearlmuter (Eds.). *The processing and acquisition of reference* (pp. 239–271). Cambridge: MIT Press.
- Hagoort, P., & Levinson, S. C. (2014). Neuropragmatics. In M. S. Gazzaniga, & G. R. Mangun (Eds.). *The cognitive neurosciences* (pp. 667–674). (5th ed.). Cambridge, Mass: MIT Press.
- Hanna, J. E., & Tanenhaus, M. K. (2004). Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science*, 28, 105–115.
- Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, 49, 43–61.
- Harrington Stack, C. M., James, A. N., & Watson, D. G. (2018). A failure to replicate rapid syntactic adaptation in comprehension. *Memory & Cognition*. <https://doi.org/10.3758/s13421-018-0808-6>.
- Hoekstra, R. A., Bartels, M., Cath, D. C., & Boomsma, D. I. (2008). Factor structure, reliability and criterion validity of the Autism-Spectrum Quotient (AQ): A study in Dutch population and patient groups. *Journal of Autism and Developmental Disorders*, 38, 1555–1566.
- Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59, 91–117.
- Huang, Y. T., & Snedeker, J. (2018). Some inferences still take time: Prosody, predictability, and the speed of scalar implicatures. *Cognitive Psychology*, 102, 105–126.
- Huetting, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137, 151–171.
- Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science*, 7, 46–50.
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, 11, 32–38.
- Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, 124, 66–71.
- Kamide, Y., Altmann, G., & Haywood, S. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133–156.
- Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of Experimental Psychology: General*, 133, 450–467.
- Kim, C. S., Gunlogson, C., Tanenhaus, M. K., & Runner, J. T. (2015). Context-driven expectations about focus alternatives. *Cognition*, 139, 28–49.
- Knoeferle, P., Crocker, M., Scheepers, C., & Pickering, M. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, 95, 95–127.
- Koornneef, A. W., & Sanders, T. J. (2013). Establishing coherence relations in discourse: The influence of implicit causality and connectives on pronoun resolution. *Language and Cognitive Processes*, 28, 1169–1206.
- Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology*, 51, 141–178.
- Kulakova, E., & Nieuwland, M. S. (2016). Pragmatic skills predict online counterfactual comprehension: Evidence from the N400. *Cognitive, Affective and Behavioral Neuroscience*, 16, 814–824.
- Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, 122, 335–342.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1–26.
- Levinson, S. C. (2000). *Presumptive meanings*. Cambridge, MA: MIT Press.
- Levinson, S. C. (1979). *Pragmatics and social deixis: Reclaiming the notion of conventional implicature* (pp. 206–223). UC Berkeley: Berkeley Linguistics Society.
- Lowder, M. W., & Ferreira, F. (2016). Prediction in the processing of repair disfluencies: Evidence from the Visual-World paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 1400–1416.
- Maschler, Y., & Schiffrin, D. (2015). Discourse markers: Language, meaning, and context. In D. Tannen, H. Hamilton, & D. Schiffrin (Eds.). *The handbook of discourse analysis* (pp. 189–221). (2nd ed.). Chichester, UK: John Wiley & Sons Ltd.
- Mortier, L., & Degand, L. (2009). Adversative discourse markers in contrast: The need for a combined corpus approach. *International Journal of Corpus Linguistics*, 14, 338–366.
- Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, 63, 324–346.
- Nieuwland, M. S. (2015). The truth before and after: Brain potentials reveal automatic activation of event knowledge during sentence comprehension. *Journal of Cognitive Neuroscience*, 27, 2215–2228.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204–238.
- Oostdijk, N. (2000). Het Corpus Gesproken Nederlands. *Nederlandse Taalkunde*, 5, 280–284.
- Politzer-Ahles, S., Xiang, M., & Almeida, D. (2017). "Before" and "after": Investigating the relationship between temporal connectives and chronological ordering using event-related potentials. *Plos ONE*, 12, e0175199.
- Potts, C. (2007). Into the conventional-implicature dimension. *Philosophy Compass*, 2, 665–679.
- Potts, C. (2015). Presupposition and implicature. In S. Lappin, & C. Fox (Eds.). *The handbook of contemporary semantic theory* (pp. 168–202). (2nd ed.). Oxford: Wiley-Blackwell.
- Rommers, J., Meyer, A. S., Praamstra, P., & Huetting, F. (2013). The contents of predictions in sentence comprehension: Activation of the shape of objects before they are referred to. *Neuropsychologia*, 51, 437–447.
- R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: < <http://www.R-project.org/> > .
- Schegloff, E. A. (2010). Some other "uh(m)"s. *Discourse and Society*, 47, 130–174.
- Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.
- Scholman, M. C. J., Rohde, H., & Demberg, V. (2017). "On the one hand" as a cue to anticipate upcoming discourse structure. *Journal of Memory and Language*, 97, 47–60.
- Sedivy, J. C. (2002). Invoking discourse-based contrast sets and resolving syntactic ambiguities. *Journal of Memory and Language*, 46, 341–370.
- Smith, S., & Jucker, A. (2002). Discourse markers as turns: Evidence for the role of interactional sequence. In A. Fetzer, & C. Meierkord (Eds.). *Rethinking sequentiality: Linguistics meets conversational interaction (Pragmatics & Beyond New Series 103)* (pp. 151–178). Amsterdam/Philadelphia: John Benjamins.
- Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science*, 268, 1632–1634.
- Tolins, J., & Fox Tree, J. E. (2016). Overhearers use addressee backchannels in dialog comprehension. *Cognitive Science*, 40, 1412–1434.
- Traugott, E. C. (2010). Revisiting subjectification and intersubjectification. In K. Davids, L. Vandelanotte, & H. Cuyckens (Eds.). *Subjectification, Intersubjectification and Grammaticalization* (pp. 29–71). Berlin: Mouton De Gruyter.
- Traugott, E. C., & Dasher, R. B. (2002). *Regularity in semantic change*. Cambridge: Cambridge University Press.
- Trude, A. M., & Brown-Schmidt, S. (2011). Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes*, 27, 979–1001.
- Verhagen, A. (2005). *Constructions of intersubjectivity. Discourse, syntax, and cognition*. Oxford: Oxford University Press.
- Wang, Y.-F., Tsai, P.-H., & Yang, Y.-T. (2010). Objectivity, subjectivity and intersubjectivity: Evidence from *qishi* ('actually') and *shishishang* ('in fact') in spoken Chinese. *Journal of Pragmatics*, 42, 705–727.
- Xiang, M., & Kuperberg, G. (2015). Reversing expectations during discourse comprehension. *Language, Cognition and Neuroscience*, 30, 648–672.